Antitoxin autoregulation of *M. tuberculosis* toxin-antitoxin expression through negative cooperativity arising from multiple inverted repeat sequences

Izaak N. Beck[a], Ben Usher[a], Hannah G. Hampton[b], Peter C. Fineran[b], Tim R. Blower[a,*]

[a]Department of Biosciences, Durham University, Stockton Road, Durham, DH1 3LE, UK

[b]Department of Microbiology and Immunology, University of Otago, PO Box 56, Dunedin 9054, New Zealand.

[*]To whom correspondence may be addressed. Email: timothy.blower@durham.ac.uk, tel: +44(0)1913343923.

ABSTRACT

Toxin-antitoxin systems play key roles in bacterial adaptation, including protection from antibiotic assault and infection by bacteriophages. The type IV toxin-antitoxin system AbiE encodes a DUF1814 nucleotidyltransferase-like toxin, and a two-domain antitoxin. In *Streptococcus agalactiae*, the antitoxin AbiEi negatively autoregulates *abiE* expression through positively co-operative binding to inverted repeats within the promoter. The human pathogen *Mycobacterium tuberculosis* encodes four DUF1814 putative toxins, two of which have antitoxins homologous to AbiEi. One such *M. tuberculosis* antitoxin, named Rv2827c, is required for growth and whilst the structure has previously been solved, the mode of regulation is unknown. To complete the gaps in our understanding, we first solved the structure of *S. agalactiae* AbiEi to 1.83 Å resolution for comparison with *M. tuberculosis* Rv2827c. AbiEi contains an N-terminal DNA binding domain and C-terminal antitoxicity domain, with bilateral faces of opposing charge. The overall AbiEi fold is similar to Rv2827c, though smaller, and with a 65^{o} difference in C-terminal domain orientation. We further demonstrate that, like AbiEi, Rv2827c can autoregulate toxin-antitoxin operon expression. In contrast to AbiEi, the $P_{rv2827c}$ promoter contains two sets of inverted repeats, which bind Rv2827c with differing affinities depending on the sequence consensus. Surprisingly, Rv2827c bound with negative co-operativity to the full $P_{rv2827c}$ promoter, demonstrating an unexpectedly complex form of transcriptional regulation.

KEYWORDS

42 Toxin-antitoxin systems, negative co-operativity, *Mycobacterium tuberculosis*, abortive infection,
43 AbiE

44

INTRODUCTION

Toxin-antitoxin (TA) systems are encoded by genetic loci that are widely distributed throughout prokaryotic genomes. They can play pivotal roles in bacterial physiology and in managing stress responses, helping bacteria to survive nutrient limitation, immune system attack, antibiotic treatment and predation by bacteriophages [1–5]. TA systems are commonly found on mobile genetic elements, contributing to the stability of plasmids, superintegrons, cryptic prophages and conjugative transposons [6–8]. The majority of TA systems encode two components, a toxic protein that generally targets essential cellular processes, and an antagonistic antitoxin [4]. This antitoxin negates toxin activity when cells are growing in favorable conditions. Under stressful conditions, the antitoxin is preferentially degraded and the toxin is released, thereby reducing growth rate as a means to survive with minimal metabolic burden until favorable conditions return [9,10]. Activation of the toxin following bacteriophage infection can also lead to the removal of the infectious bacteriophage particle from the environment, thereby providing a population level protection from viruses referred to as abortive infection (Abi) [11,12].

TA systems have been divided into six types according to the nature of the toxin and antitoxin (whether they are RNA or protein), and the mechanism of toxin antagonism [4]. Type IV systems differ from all others in that the antitoxin and toxin do not directly interact, instead, the antitoxin antagonises the activity of the toxin [13–15]. There are multiple examples wherein TA systems provide a phage-resistant Abi phenotype, although not all identified Abi systems act as *bona fide* TA systems [5,15–20]. A recently characterised Abi system, AbiE from *Streptococcus agalactiae* V/R 2603, has been shown to act as a type IV TA system [15]. AbiE encodes a DUF1814-family toxin (AbiEii), and a COG5340-family antitoxin (AbiEi) (**Fig. 1A**) [15]. The *S. agalactiae* AbiE COG5340 antitoxin will herein be referred to as AbiEi. AbiEii is a putative nucleotidyltransferase (NTase) that

specifically binds GTP [15]. This DUF1814 family is widespread, present in over 5,500 bacterial, archaeal and fungal genomes, though not all examples are genetically linked to putative antitoxins.

TA systems are remarkably abundant in *Mycobacterium tuberculosis*, which encodes more than 80 examples, and these are thought to have contributed to *M. tuberculosis* having become the most successful human pathogen [21–24]. *M. tuberculosis* H37Rv encodes four DUF1814-family NTase-like putative toxins, namely Rv0078A, Rv0836c, Rv1045 and Rv2826c (**Fig. 1A**). Akin to AbiEii from *S. agalactiae*, both Rv1045 and Rv2826c have a cognate COG5340-family antitoxin (**Fig. 1A**). Transposon mutagenesis studies have identified the cognate antitoxins of these systems (Rv1044 and Rv2827c) as essential for lab growth [25,26], suggesting that Rv1045 and Rv2826c toxins are functional in *M. tuberculosis*. The *M. tuberculosis* COG5340 proteins will herein be referred to by their respective 'Rv' identifiers, Rv1044 and Rv2827c. Characterizing and understanding the regulation of these loci is of interest for developing new therapies against the pathogen.

Autoregulation of TA system expression is a hallmark of type II TA systems and can be either positive or negative [27,28]. The antitoxin AbiEi from *S. agalactiae* has been biochemically characterized [15,29] and functions as both an antitoxin and a transcriptional repressor. That is, AbiEi negatively autoregulates *abiE* expression. Here, the gene product suppresses its own production, through positively co-operative binding of two AbiEi monomers to inverted repeats in the promoter region. Full length AbiEi is required for negative autoregulation and induced bending of the promoter DNA. We previously proposed that this bending was facilitated by the two AbiEi monomers interacting via their C-terminal domains (CTDs) [29]. In contrast to type II autoregulation, for which conditional co-operativity is observed, co-expression of the cognate toxin AbiEii does not enhance transcriptional repression [15]. We therefore sought to determine the similarities in the structure and function of

AbiEi and Rv2827c. While the structure of the *M. tuberculosis* putative antitoxin Rv2827c has been solved as part of a structural genomics initiative [30], its biological function was not explored and it has not been biochemically characterized. We present the solved structure of *S. agalactiae* AbiEi, demonstrating structural homology between the COG5340 antitoxins, and biochemically characterize the molecular interactions underpinning transcriptional repression by Rv2827c. Interestingly, this is a more complex autoregulatory system than previously seen for AbiEi [29].

MATERIALS AND METHODS

Bacterial strains and culture conditions

E. coli DH5α (Invitrogen), BL21 (DE3) (Invitrogen) and ER2566 (New England Biolabs) were routinely grown at 37 °C in Luria-Broth (LB), M9 minimal (M9M), or 2x YT media supplemented when necessary with ampicillin (Ap, 50 μg/ml), spectinomycin (Sp, 100 μg/ml), tetracycline (Tc, 10 μg/ml), isopropyl-β-D-thiogalactopyranoside (IPTG, 1 mM), L-arabinose (L-ara, 0.1% w/v) or D-glucose (glu, 0.2% w/v). Bacterial cell density was measured using a WPA Biowave C08000 at 600 nm (OD_{600}).

DNA isolation and manipulation

All oligonucleotides used in this study were obtained from Integrated DNA Technologies (**Table S1**). Plasmid and PCR-amplified DNAs were purified using Monarch kits (NEB). Digests, ligations, transformations and agarose gel electrophoresis steps were performed by standard techniques. All constructed plasmids (**Table S2**) were confirmed by sequencing using an ABI 3730 DNA sequencer and 4Peaks.

Protein expression constructs were made by Ligation Independent Cloning (LIC) [31]. Target genes were cloned into plasmid pSAT1-LIC, which generates N-terminal His_6-SUMO fusions with the target ORF. Primers TRB1048/TRB1049 were used to amplify *abiEi* from pRLD30, for LIC insertion into pSAT1-LIC, producing pTRB525. Primers TRB1022/TRB1023 were used to amplify *rv2827c* from pPF658, also for LIC insertion into pSAT1-LIC, producing pTRB493. Primers TRB1018/TRB1019 were used to amplify *rv1044* from *M. tuberculosis* H37Rv genomic DNA (ATCC), again for LIC insertion into pSAT1-LIC, producing pTRB491.

For promoter activity assays, regions upstream of *abiEi*, *rv2827c* and *rv1044* were cloned into pRW50 [32]. The 99 bp region upstream of *abiEi* was amplified from pPF680 using primers TRB1072/TRB1047, then digested with EcoRI/HindIII and ligated into pRW50 cut with the same enzymes, producing pTRB486. The 500 bp regions upstream of *rv2827c* and *rv1044* were amplified from H37Rv genomic DNA, using primers TRB1042/TRB1043 and TRB1040/TRB1041, respectively. The amplicons were digested with EcoRI/HindIII and ligated into pRW50 cut with the same enzymes, producing pTRB484 and pTRB483, respectively. Antitoxin genes *abiEi, rv2827c and rv1044* were cloned into pTA100, a pQE-80 derivative [5]. *S. agalactiae abiEi* was amplified from pRLD30 using primers TRB1052/TRB1053 the digested with EcoRI/HindIII and ligated into pTA100 cut with the same enzymes, producing pTRB481. *M. tuberculosis rv2827c* and *rv1044* were amplified from H37Rv genomic DNA, using primers PF1334/PF1335 and PF1330/PF1331 respectively. The amplicons were digested with NdeI/SpeI and ligated into pTA100 cut with the same enzymes, producing pPF658 and pPF658, respectively.

Protein expression and purification

To express AbiEi, Rv2827c and Rv1044 for crystallization and/or biochemistry, *E. coli* ER2566 (for native protein) or BL21 (DE3) (for labelled protein) were transformed with pTRB525, pTRB493 or pTRB491, respectively. For native protein, overnight cultures were re-seeded 1:100 into 2 L flasks containing 1 L 2x YT. Cells were grown at 150 rpm in 37 °C until an OD_{600} of 0.6-0.8 was reached, whereupon expression was induced by the addition of IPTG (1 mM). Cells were left to grow for 16 h at 17 °C, shaking at 150 rpm.

For incorporation of selenomethionine into AbiEi, the SeMet kit (Molecular Dimensions) was used. Starter cultures of BL21 (DE3) pTRB525, starter cultures were grown for 8 hours in LB at 37 °C with 200 rpm shaking. This culture was used to inoculate (1:500) a 50 mL overnight of Molecular Dimensions Selenomethionine Base medium supplemented with Molecular Dimensions Nutrient Mix. This overnight was then used to inoculate (1:100) 1 L of the same Base Medium with Nutrient Mix and cells were grown at 37 °C with 180 rpm shaking. At OD_{600} 0.8 cells were pelleted by centrifugation at 4200 x g, resuspended in fresh Base Medium with Nutrient Mix (Molecular Dimensions) and supplemented with an amino acid mix to promote feedback inhibition of methionine synthesis (0.1 mg/ml L-lysine hydrate, 0.1 mg/ml L-threonine, 0.1 mg/ml L-phenylalanine, 0.05 mg/ml L-leucine, 0.05 mg/ml L-isoleucine, 0.05 mg/ml L-valine). Cells were grown for a further 30 minutes at 37 °C with shaking at 180 rpm before the addition of 250x SelenoMethionine Solution (Molecular Dimensions) to a final concentration of 40 µg/mL. Cells were grown for a further 15 minutes at 37 °C with shaking at 180 rpm before antitoxin expression was induced with IPTG (1 mM), and samples were left to grow overnight at 175 rpm in 18 °C.

For native protein purification, bacteria were harvested by centrifugation at 4200 x g and the pellets were resuspended in buffer A500 (20 mM Tris-HCl pH 7.9, 500 mM NaCl, 5 mM imidazole and 10% glycerol). Cells were lysed by sonication at 40 kpsi, then centrifuged (45,000 x g, 4 °C). The clarified lysate was next passed over a HisTrap HP column (GE Healthcare), washed for ten column volumes with A500, followed by ten column volumes of buffer A100 (20 mM Tris-HCl pH 7.9, 100 mM NaCl, 5 mM imidazole and 10% w/v glycerol), then eluted directly onto a HiTrap Q HP column (GE Healthcare) with buffer B100 (20 mM Tris-HCl pH 7.9, 100 mM NaCl, 250 mM imidazole and 10% w/v glycerol). The Q HP column was transferred to an Äkta Pure (GE Healthcare), washed with three column volumes of A100, then proteins were eluted using a gradient from 100% A100 to 100% buffer C1000 (20 mM Tris-HCl pH 7.9, 1000 mM NaCl and 10% w/v glycerol). Fractions containing the

protein peak were analysed by SDS-PAGE, pooled and incubated overnight at 4 °C with hSENP2 SUMO protease to cleave the His_6-SUMO tag from the target protein. The following day, the samples were passed through a second HisTrap HP column and the flow-through fractions containing untagged target protein were collected. The same procedure was used for labelled protein, except seleno-AbiEi precipitated on column at A100, contaminants were removed with B100, and remaining folded seleno-AbiEi was eluted with B500, followed by SENP cleavage and a second HisTrap column purification. Proteins were dialysed overnight at 4 °C into buffer X (20 mM Tris-HCl pH 7.9, 200 mM NaCl and 2.5 mM DTT). Crystallization samples were concentrated, quantified and stored on ice, then either used immediately or flash-frozen in liquid N_2 for storage at -80 °C.

Protein crystallization

Native and selenomethionine-derivatized AbiEi were concentrated to 12 mg/ml in buffer X (see above). Initial native AbiEi crystallization screens were performed using commercial screens (Molecular Dimensions) set with an Innovadyne Screenmaker robot, making 200:100 nl and 100:100 nl protein:condition sitting drops at 18 ^{o}C. After initial screening and optimization, native AbiEi formed thick needles in 0.02 M Sodium/Potassium phosphate, 0.1 M Bis-Tris Propane pH 6.5, 20% PEG 3350. Selenomethionine-derivatized AbiEi crystals grew in 0.2 M Sodium acetate trihydrate, 0.1 M Bis-Tris Propane pH 6.5, 15% PEG 3350. To harvest, 20 µl of condition reservoir was added to 20 µl of glycerol and mixed quickly by vortexing; an equal volume of this mixture was then added to the drop volume. After addition of cryo buffer, crystals were immediately extracted using a nylon loop and flash-frozen in liquid N_2.

X-ray data collection and structure determination

Diffraction data were collected at Diamond Light Source on beamlines I03 (AbiEi native) and I03 (AbiEi selenomethionine-derivatized) (Table 1). Single, 360°, datasets were collected from three native AbiEi crystals and merged using iSpyB (Diamond Light Source). Two, 360°, datasets from AbiEi selenomethionine-derivatized crystals measured at the selenium peak (0.9793 Å) were also merged using iSpyB. An additional AbiEi selenomethionine-derivatized dataset was collected at selenium high remote (0.9641 Å) wavelength. Diffraction data were processed with XDS [33,34], and then AIMLESS from CCP4 [35] was used to corroborate the spacegroups (Table 1). The crystal structure of AbiEi was solved by MAD, by providing the SHELX suite [36] in CCP4 with the native, peak and high remote datasets. The solved starting model for AbiEi was built in REFMAC [37] and BUCCANEER [38]. The model was then iteratively refined and built using PHENIX [39] and COOT [40], respectively. The quality of the final model was assessed using COOT and the wwPDB validation server [41]. Structural figures were generated using PyMol (Schrödinger). Structural alignments were performed using PROMALS3D [42].

Electrophoretic Mobility Shift Assays

Conservation of IR sequences was determined using MView [43] and WebLogo [44]. Promoter region probes were amplified from synthesised templates (Table S1). Each template was made with a common downstream region, matching the initial part of *lacZ* from pRW50. For each probe template, a unique upstream forward primer was used in combination with a common reverse primer, which was either untagged or had been conjugated to a fluorescein tag for visualization (Table S1). The probes contained either the native promoter regions, or combinations of WT IR sequences and mutant sequences of polyC.

Proteins were diluted to appropriate concentrations using diluent buffers matching their storage buffer constitution. Each binding reaction contained 2 µL of 5x EMSA binding buffer (750 mM KCl, 50 mM Tris-HCl pH 8.0, 2.5 mM EDTA pH 8.0, 0.5 % Triton X-100, 1 mM DTT, 55% glycerol), 250 fmoles of fluorescently labelled probe, 0.1 µL BSA (10 mg/ml), 1 µL poly(d[IC]) (1 mg/mL), 1 µL of diluted protein or buffer control and water to a final volume of 10 µL. Native 0.5x TBE polyacrylamide gels (at either 7% or 5% acrylamide, as required) were pre-run at 150 V and 4 ^{o}C for 2 h. Binding reactions were titrated at protein concentrations from zero to an appropriate upper limit, and incubated at 20 ^{o}C for 30 min. Non-specific binding controls used an additional excess of 2.5 µM TRB1108 template DNA amplified with TRB1109 as forward primer, and non-labelled reverse primer. Specific binding controls used additional excess of 2.5 µM unlabelled specific probe DNA. Samples were then separated by native polyacrylamide gel electrophoresis at 200 V and 4 ^{o}C for 45 min.

Native polyacrylamide gels were then visualised using the Amersham Biosciences Typhoon 9400 on variable mode image in fluorescence mode, emission filter 526 SP. Sensitivity was set to normal. Band intensities were calculated using the grid scan feature and triplicate data processed in Prism (GraphPad Software). Fractional saturation curves were produced with fractional saturation, Y, varying from 0 – 1.0. Y values are calculated by (Y/(Y+(1-Y))) and plotted against protein concentration. Data were converted to the Hill plot to analyse the degree of cooperativity in the binding events, characterised by the Hill coefficient (slope of the plot at log(θ)=0). The Hill plot is constructed by plotting logθ against log[protein], with θ defined as (θ = (Y/(1-Y))). Dissociation coefficients (Kd) can also be extracted from the Hill plot as Kd = $10^{X\text{-intercept}}$. Mean and standard error of the mean values are derived from at least three independent experiments.

Promoter activity assays

Promoter regions were cloned into the promoterless *lacZ* fusion plasmid, pRW50 [32]. Antitoxin genes were cloned into the pQE-80 derivative, pTA100 [5] for tight control of antitoxin expression. Construction is detailed above. Promoter activity assays were performed as described previously [45,46]. Briefly, *E. coli* DH5α were co-transformed with the *lacZ* reporter constructs pTRB483 (P_{rv1044}), pTRB484 ($P_{rv2827c}$) or pTRB486 (P_{abiEi}), and the IPTG-inducible pTA100-antitoxin plasmids pPF656 (Rv1044), pPF658 (Rv2827c) or pTRB481 (AbiEi). Transformants were re-seeded from overnight cultures and grown in 37 °C at 200 rpm in LB supplemented with Tc, Sp, and with/without IPTG until mid-log phase, then 80 µl of cells were added to 120 µl master mix (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 36 mM β-mercaptoethanol, 166 µl/ml T7 lysozyme, 1.1 mg/ml ONPG, and 6.7 % PopCulture Reagent (Merck Millipore)) in corresponding wells of a 96-well plate. This was then placed in a SPECTROstar Nano absorbance plate reader (BMG LABTECH) set to 30 °C with shaking at 500 rpm, wherein OD_{600} and OD_{420} readings were taken every 90 seconds for 1 hour. Data analysis was performed in the MARS Data Analysis software package (BMG LABTECH). The kinetic OD_{420} readings were converted into the slope of OD_{420} over time (OD_{420}/min). These values were multiplied by 5000 and divided by the OD_{600} reading from the first time point to generate Miller Units (mU). Plotted data are the normalized mean and standard deviation obtained from three independent experiments.

Accession number

The crystal structure of AbiEi has been deposited in the Protein Data Bank under accession number 6Y8Q.

RESULTS

AbiEi-family antitoxins contain conserved structural features

We had previously hypothesized that there was structural similarity between the biochemically characterised antitoxin AbiEi from *S. agalactiae* [29] and the structurally characterised homologue, Rv2827c [30]. We sought to confirm structural and biochemical similarity between these two proteins, and within the broader COG5340 antitoxins. To begin, we solved a 1.83 Å structure of AbiEi by X-ray crystallography (**Figs. 1B and C, Table 1**). There were four copies of AbiEi in the asymmetric unit, forming minor crystal contacts that are not predicted to be biologically relevant, and each copy contains minor variations in domain orientation, indicating some flexibility. Together with previous size exclusion chromatography data [29], we concluded that AbiEi is a 23 kDa monomer in solution.

AbiEi contains an N-terminal winged helix-turn-helix DNA-binding domain and a C-terminal antitoxin domain, connected by a short linker (**Fig. 1B**). Mutagenesis studies have demonstrated that full-length AbiEi is required for negative autoregulation of the P_{abiE} promoter, whilst the C-terminal domain alone is sufficient for antitoxicity against the effects of AbiEii [15]. The N-terminal domain contains three α-helices, followed by three beta-strands forming an antiparallel sheet (**Fig. 1B**). The C-terminal domain begins with a single α-helix that is separated from a six-helix bundle by a row of four β-strands, which themselves pair into parallel and antiparallel β-sheets (**Fig. 1B**). One face of AbiEi is positively charged, and the reverse face is negatively charged (**Fig. 1C**). The positive side corresponds with the site of positively charged sidechains distributed throughout the N-terminal and C-terminal domains, which have previously been shown to be vital for DNA-binding and autoregulation through mutagenesis studies [29]. When AbiEi is compared with Rv2827c, both are monomers and it is clear that the two antitoxins share the two-domain structure and charge features (**Figs. 1B-E**).

When AbiEi and Rv2827c are aligned via the N-terminal winged helix-turn-helix domain, the respective C-terminal domains differ in position relative to the N-terminal domains by approximately 65^{o} (**Fig. 2A**). We propose that these different poses captured in the crystal structures might reflect variable positions of the C-terminal domains potentially allowed by a linker joining the two domains. The stability of the B-factors for the subdomains AbiEi and Rv2827c, alongside lack of significant variation in the domain orientations within the asymmetric unit indicates a preferred state has been captured in the crystal. This however would require further analysis in solution. The extensive nature of the AbiEi charged surface, the requirement for the full AbiEi protein for autoregulation [15], and the presence of a flexible linker altogether indicate the full protein is needed for DNA interactions and DNA bending as per our previously proposed model [29].

When the N-terminal domain helices and C-terminal domains from AbiEi and Rv2827c are separated and structurally superposed, it is possible to see an approximate overlay between corresponding regions, with RMSDs of 3.04 Å for the N-terminal helices and 3.41 Å for the C-terminal domains (**Fig. 2B and C**). The N-terminal domains have conserved positioning of key helices H2 and H3, which are used within winged helix-turn-helix domains for stabilization and DNA recognition, respectively [47] (**Fig. 2B**). The C-terminal domain of AbiEi is the smaller of the two; performing a structure-based sequence alignment of AbiEi and Rv2827c shows that Rv2827c has an extended C-terminal domain 55 amino acids longer than AbiEi (**Fig. S1**). Despite this extension, AbiEi and Rv2827c share a conserved common core fold of unknown function (**Fig. 2C**). When AbiEi was compared against the PDB to look for similar structures, using the DALI server [48], Rv2827c was the top hit followed by bacterial antibiotic-modifying adenylyltransferases (PDB codes 5KQJ, 4FO1), and a putative fungal NTase (PDB: 5UVD). These putative biochemical activities for AbiEi match well with the NTase

activity of the cognate toxin AbiEii [15]. Overall, despite differing captured poses and discrepancy in size, AbiEi and Rv2827c are markedly similar in domain structure, fold and surface charge and are therefore structural homologues.

It has been shown that the AbiEi C-terminal domain is required for negative autoregulation and likely contributes to positive cooperativity through C-terminal domain interactions [29]. The cons-PPISP server [49] was used to highlight the residues most likely to be critical for protein-protein interactions for both AbiEi and Rv2827c (**Fig. 2D-E**). In the AbiEi monomer, 16 identified residues were clustered at the C-terminus, forming a putative site for interaction (**Fig. 2D**). For Rv2827c, however, the diffuse scattering of 34 identified residues along the structure (**Fig. 2E**) predicts that there may be no obvious interface for protein-protein interactions. This is reinforced by the different positioning of the CTD seen in Rv2827c (**Fig 2A**). These findings suggest that the interactions of AbiEi C-terminal domains could contribute to positive co-operativity in promoter binding, supporting our previously proposed model, whereas for Rv2827c, such interactions are unlikely to occur and indicate a different mechanism of DNA-binding and autoregulation.

Rv2827c binds two sets of inverted repeats

AbiEi binds to two 23 bp inverted repeats (IR1 and IR2) within the promoter of P_{abiEi}, which are separated by 3 bp [29] (**Fig. 3A**). Examination of the upstream region of $P_{rv2827c}$ revealed two pairs of 23 bp inverted repeats within the region -1 to -131 bp from the *rv2827c* start codon, that also overlap the promoter (**Fig. 3A**). These four repeats (IR1 to IR4) are arranged in tandem with a 4 bp gap between the two pairs of inverted repeats and a 13 bp gap between the repeats within each pair (**Fig. 3A**). As P_{abiEi} repeats are separated by 3 bp and the repeats within pairs from $P_{rv2827c}$ are separated by 13 bp, it is possible the additional 10 bp accommodates binding of the larger C-

terminal domains of Rv2827c (**Fig. S1**). Using the bacterial promoter prediction software CNNPromoter_b [50], the IR3-IR4 repeats were predicted to straddle a binding site for the primary *M. tuberculosis* sigma factor SigA [51]. As Rv2827c binding would sterically hinder sigma factor binding, in turn, this would prevent transcription of the operon by RNA polymerase. When IR1-IR4 sequences from $P_{rv2827c}$ were aligned with IR1-IR2 sequences from P_{abiEi}, the sequence similarity indicated possible conservation of binding sequence (**Fig. 3B**). We therefore hypothesized that despite sharing low protein sequence identity (17.7%), Rv2827c might bind these $P_{rv2827c}$ inverted repeats similarly to AbiEi binding its cognate P_{abiEi} repeats.

The four $P_{rv2827c}$ 23 bp inverted repeats were first tested as two consecutive pairs, to allow a direct comparison to the arrangement of P_{abiEi} [29]. Analysis began with IR3-IR4, the pair overlapping the transcriptional start and therefore analogous to IR1-IR2 of P_{abiEi} (**Fig. 4A**). Using electrophoretic mobility shift assays (EMSAs), Rv2827c was shown to bind both of the IR3-IR4 inverted repeats within the -1 to -71 region (**Fig. 4B**). Sequential removal of the inverted repeats by mutating one, the other or both to poly-C tracts reduced Rv2827c-DNA interaction to a single binding event (**Fig. 4C and D**) or ablated binding completely (**Fig. 4E**). Analysis of IR3-IR4 binding (**Fig. 4B and F**) showed weak saturation of binding. The calculated Hill co-efficient indicates that IR3-IR4 binding by Rv2827c is not co-operative (**Fig. 4G**).

Similar results were obtained when testing the IR1-IR2 repeats within the -61 to -131 region of $P_{rv2827c}$ (**Fig. 5A-G**). In this case, there was greater saturation of binding to IR1-IR2 (**Fig. 5F**) than to IR3-IR4 (**Fig. 4F**). The Hill co-efficient surprisingly indicated weakly negative co-operativity in Rv2827c binding to IR1 and IR2 (**Fig. 5G**), in comparison to the non-co-operative binding observed with IR3 and IR4 (**Fig. 4G**). To allow direct comparison between model systems, we also performed the same

assays with purified AbiEi and probes for P_{abiEi} (**Fig. S2**). This corroborated previous data [29] and under our experimental conditions, AbiEi bound more tightly to its cognate inverted repeats (**Fig. S2F**), than either Rv2827c binding to IR3-IR4 (**Fig. 4F**) or IR1-IR2 (**Fig. 5F**), and also demonstrated clear positive co-operativity (**Fig. S2G**).

Due to similarity of structure, functionality and cognate DNA inverted repeat sequences, we hypothesized that AbiEi and Rv2827c might bind their respective non-cognate promoter regions. However, AbiEi did not bind either pair of inverted repeats from $P_{rv2827c}$ (**Fig. S3A and B**). Similarly, Rv2827c did not bind IR1-IR2 of P_{abiEi} (**Fig. S3C**).

Rv2827c binds with negative co-operativity

Having investigated the two sets of $P_{rv2827c}$ inverted repeats independently, a full-length probe covering the $P_{rv2827c}$ region -1 to -131 was generated to examine the interaction of Rv2827c protein with all four inverted repeats. Using EMSAs, four distinct protein-bound DNA species were observed, indicating that all four inverted repeats can be bound simultaneously by Rv2827c (**Fig. 6A**). The four binding sites did not fully saturate (**Fig. 6B**), and the Hill coefficient confirmed negatively co-operative binding of Rv2827c across these four inverted repeats (**Fig. 6C**). Displaying the saturation curve data on a semi-log scale highlights breaks and multiple distinct gradients in the binding curve, eluding to multiple individual binding events (**Fig. 6D**). Negatively co-operative binding by Rv2827c to $P_{rv2827c}$ contrasts with the positive co-operativity observed for AbiEi binding to P_{abiEi} [29].

Our earlier data using mutant probes provided insight into how Rv2827c binds to individual repeats (**Fig. 4C and D, Fig. 5C and D**). These were used to calculate the binding affinity of Rv2827c for each

individual IR sequence, with IR1 most tightly bound (K_d of 0.0205 µM), closely followed by IR4 (K_d of 0.121 µM), then IR2 (K_d of 0.862 µM), and finally IR3 (K_d of 11.0 µM) (**Fig. 6E-M**). This descending affinity series creates a wide range of concentrations across which negative autoregulation by Rv2827c can occur. These data demonstrate the same core principles of promoter binding are used by both AbiEi and Rv2827c, but that these have been employed evolutionarily for differing modes of regulation.

Rv1044 is a DNA-binding protein, but fails to recognize the cognate promoter

Whilst it had not been possible to identify inverted repeats within the *rv1044-rv1045* locus, two, slightly overlapping, 70 bp probes were generated to cover the 131 bp region upstream of the *rv1044* translational start site, and used to test Rv1044 binding (**Fig. S4A and B**). No binding event was observed with either probe (**Fig. S4A and B**). Nevertheless, we wanted to test whether Rv1044 was competent for DNA-binding and so cross-reacted Rv1044 with the two probes covering IR3-IR4 and IR1-IR2 of $P_{rv2827c}$, and the probe containing IR1-IR2 of P_{abiEi} (**Fig. S4C, D and E**). No binding was observed for either of the $P_{rv2827c}$ probes (**Fig. S4C and D**), but curiously, Rv1044 bound the inverted repeats of P_{abiEi} (**Fig. S4E**). Rv1044 bound more weakly than AbiEi to P_{abiEi} IR1-IR2 (**Fig. S4F**), and showed no co-operativity (**Fig. S4G**). This demonstrates that Rv1044 can bind DNA in a sequence-specific manner, and so we looked for potential targets in the *M. tuberculosis* H37Rv genome. The *abiE* IR sequences align with numerous positions in the *M. tuberculosis* genome but not upstream of any of the DUF1814 TA systems. This may indicate a potential role for Rv1044 in regulating genes outside of the *rv1044-rv1045* operon, as has been shown in other TA systems whereby antitoxins influence gene expression in biofilm formation pathways [52–54]. A further study will be needed to fully explore any potential regulatory role of Rv1044.

Rv2827c negatively autoregulates Rv2827c-Rv2826c expression

Having shown a structural similarity between the two antitoxins, we next sought to test whether the COG5340 proteins from *M. tuberculosis* could function as autoregulators, like characterized AbiEi [29]. AbiEi negatively autoregulates expression from the P_{abiEi} promoter [29]. To examine whether Rv2827c and also the second *M. tuberculosis* COG5340 protein, Rv1044, also perform negative autoregulation, we first cloned the 500 bp region upstream of each respective translational start site into a promoterless *lacZ*-reporter plasmid [32]. For comparison, the equivalent P_{abiEi}–reporter, containing the previously identified promoter region identified in the upstream 99 bp [15,29] was also tested. Both P_{abiEi} and $P_{rv2827c}$ reporters yielded expression of LacZ, but P_{rv1044} did not (**Fig. 7A**). The two active reporter constructs P_{abiEi} and $P_{rv2827c}$, were then paired with inducible plasmids for expression of the cognate antitoxins, and LacZ levels were determined with and without antitoxin induction (**Fig. 7B**). When compared to the uninduced controls, both antitoxins negatively autoregulated expression from their cognate promoters (**Fig. 7B**) demonstrating that Rv2827c and AbiEi share not only a common structure, but also a common negative autoregulatory function.

DISCUSSION

In this study we present the crystal structure of *S. agalactiae* AbiEi, which was the first type IV TA system antitoxin shown to be capable of transcriptional autoregulation through promoter binding [29]. Further to this, we have demonstrated the autoregulatory capacity of the related Rv2827c antitoxin, a protein of known structure [30]. Whilst AbiEi is a structural homologue of the Rv2827c antitoxin, and both share similar promoter architectures, they have distinct differences in their size and captured domain orientations. We also show that negative autoregulation of the $P_{rv2827c}$ promoter operates via negatively co-operative interactions.

Despite the low shared sequence similarity seen for the COG5349 antitoxins investigated (AbiEi and Rv2827c – 17.7 %; AbiEi and Rv1044 – 21.2 %; Rv2827c and Rv1044 – 24.5 %), we have demonstrated structural conservation across species. As sequences diverge, structure is conserved (**Fig. 1, Fig. 2**), which maintains the shared functionality of these antitoxins, for instance, DNA-binding (**Figs. 4–7, Figs. S2–S4**). Interestingly, sequence variation of the NTD, alongside differing promoter architectures, has resulted in at least three variations of antitoxicity. AbiEi and Rv2827c both autoregulate their own operons, albeit with contrasting types of cooperativity. Rv1044, however, may regulate genes elsewhere in the *M. tuberculosis* genome, given the lack of affinity to the *rv1044* upstream region tested (**Fig. S4A-B**) and absence of identifiable inverted repeats, but apparent DNA-binding capabilities (**Fig. S4E-G**). Further analysis will be required to identify a functional promoter for the *rv1044-rv1045* operon and confirm any potential regulatory function of Rv1044. The antitoxic CTDs have a common core fold that are predicted to have NTase activity based on structure-based functional searches [30]. Therefore, the antitoxic mechanism is likely conserved, despite low sequence similarity within these domains (**Fig. S1**). As protein sequences will be tuned to the needs of the organism, we have shown a correspondingly differential pattern of residues for

protein-protein interactions (**Fig. 2D-E**) which, alongside the different CTD positions captured (**Fig. 2A),** may contribute to individual autoregulation requirements. Our previous model predicted AbiEi C-terminal domain interactions promote positive co-operative binding and result in DNA-bending [29], however this does not appear to apply to Rv2827c. Our proposed model (**Fig. 8**) implies a possible lack of protein-protein interactions supported by predicted interaction interfaces (**Fig. 2D-E**), while not ruling out the potential for steric restriction. Rather, promoter inverted repeat sequence 'tuning' (**Fig. 3**) contributes to the negatively co-operative interaction via descending affinities.

Promoters of *M. tuberculosis* are known to be more complex than those of *E. coli*; they can stretch to 2000 bp from the start site and lack canonical elements such as the conserved -35 sequence [55–57]. Transcriptional regulation is complicated further when considering the vast number of sigma factors [58] and environmentally responsive transcription factors [59] present in *M. tuberculosis*, allowing for greater promoter sequence variation. The -10 sequence for *rv2827c-rv2826c* is a predicted recognition site for principle *M. tuberculosis* sigma factor SigA, which is usually maintained at a constant level for cellular "housekeeping" [51]. SigA also has a role in host-pathogen interactions, controlling growth rates during macrophage infection [60] and regulating virulence genes through both constitutive and upregulated expression [61–63]. Deletions of *rv2827c* cause a growth defect [25,26], suggesting SigA drives expression and that there is potential for output to be tuned by SigA and Rv2827c levels according to environmental cues. Previous reports on the type IV antitoxin CbeA demonstrate a positive effect on cytoskeletal bundling alongside antitoxicity and the ability to counteract chemical inhibitors of cytoskeletal polymerisation [13]. One study has shown Rv2827c upregulation in response to isoniazid and rifampicin treatment, albeit as part of more general TA system upregulation [64].

The IR conservation between P_{abiEi} and $P_{rv2827c}$ (**Fig. 3**) suggested that autoregulation may also occur in a biochemically comparable manner between the two. However, Rv2827c bound the pairs of inverted repeats with either no co-operativity (**Fig. 4**), or weakly negative co-operativity (**Fig. 5**). There was clear negative co-operativity when all four inverted repeats were tested (**Fig. 6**). Analysing each inverted repeat independently by mutational studies identified significant differences between the Rv2827c-IR dissociation constants (**Fig. 6E-M**). These data have allowed us to propose a model for the regulation of *rv2827c-rv2826c* (**Fig. 8**). As *rv2827c* is needed for normal growth, this suggests that *rv2826c* encodes a toxin capable of causing growth defects [25,65], which is antagonised by Rv2827c (**Fig. 8A**). Expression of *rv2827c-rv2826c* is negatively autoregulated by Rv2827c, and this is made possible by sequential binding of Rv2827c to the four IR sequences, in order as determined by binding affinity (**Fig. 8B**). Given the high concentration of Rv2827c required to saturate the lower affinity site IR3 (**Fig. 6J, K; Fig. 8B**), mimicking the mutational analysis performed here in promoter activity studies would provide useful insight into the function of each IR sequence. These binding events have apparent negative co-operativity, likely due to the variations in IR sequences creating a series of binding steps with ever-decreasing affinity. In order to better understand these negatively co-operative interactions further experiments are required, exploring the role of the Rv2827c CTD and increased inverted repeat spacers, akin to previous work on AbiEi [29].

Negative cooperativity was an unexpected result given the structural similarities between the N-terminal domains of AbiEi and Rv2827c (**Fig. 2B**), and the similarities of their respective promoter architectures (**Fig. 3**). Examples of negative and positive co-operativity have been found in equal abundance across all organisms [66,67]. Positive co-operativity leads to rapid saturation at a defined, short range of concentrations as seen for *abiE* [29]. In contrast, negative co-operativity of Rv2827c

binding would be expected to generate a relatively delayed response, working across a greater range of Rv2827c concentrations [67–69]. This variability in tuning according to concentration could in turn relate to the relative potency of the toxins and dosage required to have an effect in their cognate hosts. This variation is evident when comparing saturation curves of AbiEi and Rv2827c to their cognate full-length promoters (**Fig. 6B and Fig. S2F**). Compared to positive co-operativity, there is relatively little information on the presence of negatively co-operative TA-promoter interactions. However, clear evidence supports weaker binding of un-complexed type II antitoxins [52,70] when compared to the conditionally co-operative binding of TA complexes [28,52,71,72]. It is noteworthy that unlike many type II antitoxins, AbiEi and Rv2827c are fully folded and stable, and also no conditionally co-operative response was seen for AbiE, and so the conditional model proposed for many type II systems likely does not apply [15].

This study has shown that the similar structures and promoter architectures between AbiEi, Rv2827c (and indeed Rv1044) have been co-opted to form different regulatory modules. A greater understanding of how these nuances of regulation are applied in the cognate hosts may provide greater insight into the control of bacterial growth. Understanding these systems and how they regulate bacterial behaviour is thereby an important step in developing a means to control TA systems towards utilising them for their potential therapeutic value.

ACKNOWLEDGEMENTS

We thank Koen Semeijn and Ron Dy for initial plasmid construction and preliminary testing.

AUTHOR CONTRIBUTIONS

Conceptualization, all authors; Investigation, I.N.B., B.U., H.H.; Writing, all authors; Funding acquisition, P.C.F., T.R.B.; Supervision, P.C.F., T.R.B..

FUNDING

This work was supported by a Springboard Award from the Academy of Medical Sciences (SBF002\1104) [I.N.B., B.U., T.R.B.], a BBSRC NLD Doctoral Training Partnership studentship [I.N.B.], a University of Otago Doctoral Scholarship [H.G.H], a University of Otago Research Grant [P.C.F], School of Biomedical Sciences Bequest Funds, University of Otago [P.C.F] and the Matariki Network of Universities (MNU) [P.C.F].

CONFLICT OF INTEREST

The authors declare no conflict of interest.

549

550 # REFERENCES

551 [1] J.P. Norton, M.A. Mulvey, Toxin-Antitoxin Systems Are Important for Niche-Specific
552 Colonization and Stress Resistance of Uropathogenic *Escherichia coli*, PLoS Pathog. 8 (2012)
553 e1002954.

554 [2] H. Van Acker, T. Coenye, The Role of Reactive Oxygen Species in Antibiotic-Mediated Killing of
555 Bacteria, Trends Microbiol. 25 (2017) 456–466.

556 [3] S. Helaine, A.M. Cheverton, K.G. Watson, L.M. Faure, S.A. Matthews, D.W. Holden,
557 Internalization of salmonella by macrophages induces formation of nonreplicating persisters,
558 Science. 343 (2014) 204–208.

559 [4] R. Page, W. Peti, Toxin-antitoxin systems in bacterial growth arrest and persistence, Nat.
560 Chem. Biol. (2016).

561 [5] P.C. Fineran, T.R. Blower, I.J. Foulds, D.P. Humphreys, K.S. Lilley, G.P. Salmond, The phage
562 abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair, Proc. Natl.
563 Acad. Sci. U. S. A. 106 (2009) 894–899.

564 [6] N. Fraikin, F. Goormaghtigh, L. Van Melderen, Type II toxin-antitoxin systems: evolution and
565 revolutions., J. Bacteriol. (2020).

566 [7] S. Szekeres, M. Dauti, C. Wilde, D. Mazel, D.A. Rowe-Magnus, Chromosomal toxin-antitoxin
567 loci can diminish large-scale genome reductions in the absence of selection, Mol. Microbiol.
568 63 (2007) 1588–1605.

569 [8] R.A.F. Wozniak, M.K. Waldor, Integrative and conjugative elements: mosaic mobile genetic
570 elements enabling dynamic lateral gene flow, Nat. Rev. Microbiol. 8 (2010) 552–563.

[9] H. Deter, R. Jensen, W. Mather, N. Butzin, Mechanisms for Differential Protein Production in Toxin–Antitoxin Systems, Toxins. 9 (2017) 211.

[10] A.M. Hall, B. Gollan, S. Helaine, Toxin–antitoxin systems: reversible toxicity, Curr. Opin. Microbiol. 36 (2017) 102–110.

[11] S.J. Labrie, J.E. Samson, S. Moineau, Bacteriophage resistance mechanisms, Nat. Rev. Microbiol. 8 (2010) 317–327.

[12] M.C. Chopin, A. Chopin, E. Bidnenko, Phage abortive infection in lactococci: variations on a theme, Curr. Opin. Microbiol. 8 (2005) 473–479.

[13] H. Masuda, Q. Tan, N. Awano, K.-P. Wu, M. Inouye, YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in *Escherichia coli*, Mol. Microbiol. 84 (2012) 979–989.

[14] H. Masuda, Q. Tan, N. Awano, Y. Yamaguchi, M. Inouye, A novel membrane-bound toxin for cell division, CptA (YgfX), inhibits polymerization of cytoskeleton proteins, FtsZ and MreB, in *Escherichia coli*, FEMS Microbiol. Lett. 328 (2012) 174–81.

[15] R.L. Dy, R. Przybilski, K. Semeijn, G.P.C. Salmond, P.C. Fineran, A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism., Nucleic Acids Res. 42 (2014) 4590–605.

[16] T.R. Blower, X.Y. Pei, F.L. Short, P.C. Fineran, D.P. Humphreys, B.F. Luisi, G.P.C. Salmond, A processed noncoding RNA regulates an altruistic bacterial antiviral system, Nat. Struct. Mol. Biol. 18 (2011) 185–190.

[17] E. Emond, E. Dion, S.A. Walker, E.R. Vedamuthu, J.K. Kondo, S. Moineau, AbiQ, an abortive infection mechanism from *Lactococcus lactis*, Appl. Environ. Microbiol. 64 (1998) 4748–56.

593 [18] J.E. Samson, S. Spinelli, C. Cambillau, S. Moineau, Structure and activity of AbiQ, a lactococcal
594 endoribonuclease belonging to the type III toxin-antitoxin system, Mol. Microbiol. 87 (2013)
595 756–768.

596 [19] D.C. Pecota, T.K. Wood, Exclusion of T4 phage by the hok/sok killer locus from plasmid R1, J.
597 Bacteriol. 178 (1996) 2044–2050.

598 [20] R. Hazan, H. Engelberg-Kulka, *Escherichia coli* mazEF-mediated cell death as a defense
599 mechanism that inhibits the spread of phage P1, Mol. Genet. Genomics. 272 (2004) 227–234.

600 [21] I. Keren, S. Minami, E. Rubin, K. Lewis, Characterization and transcriptome analysis of
601 *Mycobacterium tuberculosis* persisters., MBio. 2 (2011) e00100-11.

602 [22] A. Sala, P. Bordes, P. Genevaux, Multiple toxin-antitoxin systems in *Mycobacterium*
603 *tuberculosis*, Toxins 6 (2014) 1002-1020.

604 [23] R.A. Slayden, C.C. Dawson, J.E. Cummings, Toxin–antitoxin systems and regulatory
605 mechanisms in *Mycobacterium tuberculosis*, Pathog. Dis. 76 (2018).

606 [24] H. Akarsu, P. Bordes, M. Mansour, D.J. Bigot, P. Genevaux, L. Falquet, TASmania: A bacterial
607 Toxin-Antitoxin Systems database, PLoS Comput. Biol. 15 (2019) e1006946.

608 [25] C.M. Sassetti, D.H. Boyd, E.J. Rubin, Genes required for mycobacterial growth defined by high
609 density mutagenesis, Mol. Microbiol. 48 (2003) 77–84.

610 [26] M.A. Dejesus, E.R. Gerrick, W. Xu, S.W. Park, J.E. Long, C.C. Boutte, E.J. Rubin, D.
611 Schnappinger, S. Ehrt, S.M. Fortune, C.M. Sassetti, T.R. Ioerger, Comprehensive essentiality
612 analysis of the *Mycobacterium tuberculosis* genome via saturating transposon mutagenesis,
613 MBio. 8 (2017) e02133-16.

614 [27] K. Gerdes, S.K. Christensen, A. Lobner-Olesen, Prokaryotic toxin-antitoxin stress response loci,

615 Nat. Rev. Microbiol. 3 (2005) 371–382.

616 [28] M. Overgaard, J. Borch, M.G. Jørgensen, K. Gerdes, Messenger RNA interferase RelE controls
617 *relBE* transcription by conditional cooperativity, Mol. Microbiol. 69 (2008) 841–857.

618 [29] H.G. Hampton, S.A. Jackson, R.D. Fagerlund, A.I.M. Vogel, R.L. Dy, T.R. Blower, P.C. Fineran,
619 AbiEi Binds Cooperatively to the Type IV abiE Toxin–Antitoxin Operator Via a Positively-
620 Charged Surface and Causes DNA Bending and Negative Autoregulation, J. Mol. Biol. 430
621 (2018) 1141–1156.

622 [30] R. Janowski, S. Panjikar, A.N. Eddine, S.H.E. Kaufmann, M.S. Weiss, Structural analysis reveals
623 DNA binding properties of Rv2827c, a hypothetical protein from *Mycobacterium tuberculosis*,
624 J. Struct. Funct. Genomics. 10 (2009) 137–150.

625 [31] C. Aslanidis, P.J. de Jong, Ligation-independent cloning of PCR products (LIC-PCR), Nucleic
626 Acids Res. 18 (1990) 6069–74.

627 [32] J. Lodge, J. Fear, S. Busby, P. Gunasekaran, N.R. Kamini, Broad host range plasmids carrying
628 the *Escherichia coli* lactose and galactose operons, FEMS Microbiol Lett. 95 (1992) 271–276.

629 [33] W. Kabsch, XDS., Acta Crystallogr. D. Biol. Crystallogr. 66 (2010) 125–32.

630 [34] W. Kabsch, Integration, scaling, space-group assignment and post-refinement, Acta
631 Crystallogr. D. Biol. Crystallogr. 66 (2010) 133–44.

632 [35] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B.
633 Krissinel, A.G.W. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A.
634 Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP4 suite and
635 current developments, Acta Crystallogr. D. Biol. Crystallogr. 67 (2011) 235–42.

636 [36] G.M. Sheldrick, IUCr, A short history of *SHELX*, Acta Crystallogr. Sect. A Found. Crystallogr. 64

(2008) 112–122.

[37] A.A. Vagin, R.A. Steiner, A.A. Lebedev, L. Potterton, S. McNicholas, F. Long, G.N. Murshudov, *REFMAC* 5 dictionary: organization of prior chemical knowledge and guidelines for its use, Acta Crystallogr. Sect. D Biol. Crystallogr. 60 (2004) 2184–2195.

[38] K. Cowtan, IUCr, The Buccaneer software for automated model building. 1. Tracing protein chains, Acta Crystallogr. Sect. D Biol. Crystallogr. 62 (2006) 1002–1011.

[39] P.D. Adams, P. V Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr. D. Biol. Crystallogr. 66 (2010) 213–21.

[40] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr D Biol Crystallogr. 60 (2004) 2126–2132.

[41] S. Gore, S. Velankar, G.J. Kleywegt, Implementing an X-ray validation pipeline for the Protein Data Bank, Acta Crystallogr. Sect. D Biol. Crystallogr. 68 (2012) 478–483.

[42] J. Pei, B.-H. Kim, N. V. Grishin, PROMALS3D: a tool for multiple protein sequence and structure alignments, Nucleic Acids Res. 36 (2008) 2295–2300.

[43] F. Madeira, Y. mi Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A.R.N. Tivey, S.C. Potter, R.D. Finn, R. Lopez, The EMBL-EBI search and sequence analysis tools APIs in 2019, Nucleic Acids Res. 47 (2019) W636–W641.

[44] G.E. Crooks, G. Hon, J.-M. Chandonia, S.E. Brenner, WebLogo: A Sequence Logo Generator, Genome Res. 14 (2004) 1188-1190.

[45] J. Schaefer, G. Jovanovic, I. Kotta-Loizou, M. Buck, A data comparison between a traditional and the single-step β-galactosidase assay, Data Br. 8 (2016) 350–352.

[46] J. Schaefer, G. Jovanovic, I. Kotta-Loizou, M. Buck, Single-step method for β-galactosidase assays in *Escherichia coli* using a 96-well microplate reader, Anal. Biochem. 503 (2016) 56–57.

[47] K.S. Gajiwala, S.K. Burley, Winged helix proteins, Curr. Opin. Struct. Biol. 10 (2000) 110–116.

[48] L. Holm, C. Sander, Protein structure comparison by alignment of distance matrices, J. Mol. Biol. 233 (1993) 123–138.

[49] H. Chen, H.-X. Zhou, Prediction of interface residues in protein-protein complexes by a consensus neural network method: Test against NMR data, Proteins Struct. Funct. Bioinforma. 61 (2005) 21–35.

[50] R.K. Umarov, V. V. Solovyev, Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks, PLoS One. 12 (2017) e0171410.

[51] S. Rodrigue, J. Brodeur, P.-E. Jacques, A.L. Gervais, R. Brzezinski, L. Gaudreau, Identification of mycobacterial sigma factor binding sites by chromatin immunoprecipitation assays, J. Bacteriol. 189 (2007) 1505–13.

[52] M. V. Merfa, B. Niza, M.A. Takita, A.A. De Souza, The MqsRA Toxin-Antitoxin System from *Xylella fastidiosa* Plays a Key Role in Bacterial Fitness, Pathogenicity, and Persister Cell Formation, Front. Microbiol. 7 (2016) 904.

[53] V.W.C. Soo, T.K. Wood, Antitoxin MqsA Represses Curli Formation Through the Master Biofilm Regulator CsgD, Sci. Rep. 3 (2013) 3186.

[54] X. Wang, T.K. Wood, Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response, Appl. Env. Microbiol. 77 (2011) 5577–5583.

[55] N. Agarwal, A.K. Tyagi, Mycobacterial transcriptional signals: requirements for recognition by RNA polymerase and optimal transcriptional activity, Nucleic Acids Res. 34 (2006) 4245–57.

[56] S.S. Shell, J. Wang, P. Lapierre, M. Mir, M.R. Chase, M.M. Pyle, R. Gawande, R. Ahmad, D.A. Sarracino, T.R. Ioerger, S.M. Fortune, K.M. Derbyshire, J.T. Wade, T.A. Gray, Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape, PLoS Genet. 11 (2015) e1005641.

[57] T. Cortes, O.T. Schubert, G. Rose, K.B. Arnvig, I. Comas, R. Aebersold, D.B. Young, Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*, Cell Rep. 5 (2013) 1121–31.

[58] P. Sachdeva, R. Misra, A.K. Tyagi, Y. Singh, The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators, FEBS J. 277 (2010) 605–626.

[59] T.R. Rustad, K.J. Minch, S. Ma, J.K. Winkler, S. Hobbs, M. Hickey, W. Brabant, S. Turkarslan, N.D. Price, N.S. Baliga, D.R. Sherman, Mapping and manipulating the *Mycobacterium tuberculosis* transcriptome using a transcription factor overexpression-derived regulatory network, Genome Biol. 15 (2014) 502.

[60] S. Wu, S.T. Howard, D.L. Lakey, A. Kipnis, B. Samten, H. Safi, V. Gruppo, B. Wizel, H. Shams, R.J. Basaraba, I.M. Orme, P.F. Barnes, The principal sigma factor sigA mediates enhanced growth of *Mycobacterium tuberculosis in vivo*, Mol. Microbiol. 51 (2004) 1551–1562.

[61] A.J. Vallecillo, C. Espitia, Expression of Mycobacterium tuberculosis pe_pgrs33 is repressed during stationary phase and stress conditions, and its transcription is mediated by sigma factor A, Microb. Pathog. 46 (2009) 119–127.

[62] G. Bagchi, S. Chauhan, D. Sharma, J.S. Tyagi, Transcription and autoregulation of the Rv3134c-devR-devS operon of *Mycobacterium tuberculosis*, Microbiology. 151 (2005) 4045–4053.

[63] J. Rengarajan, B.R. Bloom, E.J. Rubin, Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 8327–8332.

[64] A. Gupta, B. Venkataraman, M. Vasudevan, K. Gopinath Bankar, Co-expression network analysis of toxin-antitoxin loci in *Mycobacterium tuberculosis* reveals key modulators of cellular stress, Sci. Rep. 7 (2017) 5868.

[65] J.E. Griffin, J.D. Gawronski, M.A. Dejesus, T.R. Ioerger, B.J. Akerley, C.M. Sassetti, High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism, PLoS Pathog. 7 (2011) e1002251.

[66] D.E. Koshland, K. Hamadani, Proteomics and Models for Enzyme Cooperativity, J. Biol. Chem. 277 (2002) 46841–46844.

[67] A. Levitzki, D.E. Koshland, Negative cooperativity in regulatory enzymes, Proc. Natl. Acad. Sci. U. S. A. 62 (1969) 1121–1128.

[68] P. De Meyts, J. Roth, Cooperativity in ligand binding: A new graphic analysis, Biochem. Biophys. Res. Commun. 66 (1975) 1118–1126.

[69] S.H. Ha, J.E. Ferrell, Jr., Thresholds and ultrasensitivity from negative cooperativity, Science. 352 (2016) 990–993.

[70] M. Overgaard, J. Borch, K. Gerdes, RelB and RelE of Escherichia coli Form a Tight Complex That Represses Transcription via the Ribbon–Helix–Helix Motif in RelB, J. Mol. Biol. 394 (2009) 183–196.

[71] I. Cataudella, A. Trusina, K. Sneppen, K. Gerdes, N. Mitarai, Conditional cooperativity in toxin–antitoxin regulation prevents random toxin activation and promotes fast translational

recovery, Nucleic Acids Res. 40 (2012) 6424–6434.

[72] A. Garcia-Pino, S. Balasubramanian, L. Wyns, E. Gazit, H. De Greve, R.D. Magnuson, D. Charlier, N.A. van Nuland, R. Loris, Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity, Cell. 142 (2010) 101–111.

FIGURES

Fig. 1. Antitoxin AbiEi is a two-domain protein with bilateral opposingly-charged faces. (A) Scaled representation of the four *M. tuberculosis* TA systems containing NTase-like toxin genes and AbiE from *S. agalactiae*. Numbers in parentheses indicate amino acid length. All five toxins are DUF1814 proteins; Rv1044, Rv2827c and AbiEi are COG5340-containing antitoxins. Putative antitoxin Rv0837c is a COG4861 protein and the significantly shorter putative antitoxin Rv0078B is unclassified. The four *M. tuberculosis* systems were re-named as shown. (B) AbiEi antitoxin structure shown in pink cartoon representation, in two views rotated 180°. (C) Electrostatic potential of AbiEi, posed as per (B), with electropositive charge in blue and electronegative charge in red. (D) Previously solved Rv2827c structure shown in blue cartoon representation, in two views rotated 180° (PDB: 1ZEL). (E) Electrostatic potential of Rv2827c, posed as per (D), colored as per (C).

Fig. 2. AbiEi and Rv2827c are structurally similar, but have been captured in different positions with differing predicted protein interaction interfaces. (A) AbiEi (pink) and Rv2827c (blue) in cartoon representation, aligned via the N-terminal winged helix-turn-helix domains, shown as two orthogonal views. The positions of the C-terminal domains diverge at a 65^{o} angle. (B) Close-up structural superposition of the isolated N-terminal helices of AbiEi and Rv2827c, colored as per (A). The three helices (H1-3) of the N-terminal winged helix-turn-helix domains align well. (C) Close-up structural superposition of the isolated C-terminal domains of AbiEi and Rv2827c, colored as per (A). The core secondary structural features of the C-terminal domains approximate to the same positions, but the Rv2827c C-terminal domain has additional features at the C-terminus. (D) AbiEi has C-terminal residues predicted to be involved in making protein-protein interactions, which might allow positive co-operativity in AbiEi monomer binding. AbiEi is in pink cartoon representation with identified interacting residues in red, and is shown in orthogonal views. (E) Rv2827c does not have an equivalent patch of C-terminal interacting residues. Rv2827c is in blue cartoon representation, with identified interacting residues in cyan, and is shown in 180° rotation. Residues were identified using the cons-PPISP server. Rv2827c PDB code: 1ZEL.

Fig. 3. The *rv2827c-rv2826c* promoter has similar features but is more complex than the *abiE* promoter. (A) Cartoon of the *abiE* and *rv2827c-rv2826c* promoters (pink and blue, respectively), showing the relative positions of the 23 bp inverted repeats (IRs). Putative transcriptional -35, -10 and start sites, along with ribosome binding sites (RBS), are indicated where possible. (B) Alignment of the six, 23 bp, IR sequences shows consensus sequences between the *abiE* and *rv2827c-rv2826c* promoters. The alignment was made using MView and the consensus was made using WebLogo.

Fig. 4. Rv2827c binds non-co-operatively to the IR3-IR4 region of the *rv2827c-rv2826c* promoter. (A) Sequence level cartoon of the fluorescently labelled probe containing IR3-IR4, with -35, -10 and transcriptional start indicated. (B) Electrophoretic mobility shift assay (EMSA) of titrated Rv2827c with the probe in (A). (C) EMSA of titrated Rv2827c with the probe in (A) altered by replacing IR4 with polyC. (D) EMSA of titrated Rv2827c with the probe in (A) altered by replacing IR3 with polyC. (E) EMSA of titrated Rv2827c with the probe in (A) altered by replacing both IR3 and IR4

with polyC. For (B-E); protein concentrations are shown on each panel together with the binding events (0, 1 or 2); S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe; numbering -1 to -71 denotes the promoter region included in the probe, upstream of the translational start site in order to include all of IR4. (F) Fractional saturation curve plotted using the EMSA data of (B). (G) Hill plot using the EMSA data from (B). For (F) and (G), points are plotted from triplicate data and display mean values with standard error of the mean.

Fig. 5. Rv2827c binds with weak negative co-operativity to the IR1-IR2 region of the *rv2827c-rv2826c* promoter. (A) Sequence level cartoon of the fluorescently labelled probe containing IR1-IR2. (B) Electrophoretic mobility shift assay (EMSA) of titrated Rv2827c with the probe in (A). (C) EMSA of titrated Rv2827c with the probe in (A) altered by replacing IR2 with polyC. (D) EMSA of titrated Rv2827c with the probe in (A) altered by replacing IR1 with polyC. (E) EMSA of titrated Rv2827c with the probe in (A) altered by replacing both IR1 and IR2 with polyC. For (B-E); protein concentrations are shown on each panel together with the binding events (0, 1 or 2); S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe; numbering -60 to -131 denotes the promoter region included in the probe. (F) Fractional saturation curve plotted using the EMSA data of (B). (G) Hill plot using the EMSA data from (B). For (F) and (G), points are plotted from triplicate data and display mean values with standard error of the mean.

Fig. 6. Rv2827c binds with negative co-operativity to the full *rv2827c-rv2826c* promoter. (A) EMSA of titrated Rv2827c with a probe covering from -1 to -131 of the r*v2827c-rv2826c* promoter, covering IR1 to IR4. The titration was performed across two EMSA gels, with an additional zero protein lane included in the second gel for normalization. Protein concentrations are shown below each gel together with the binding events (0, 1, 2, 3 or 4); S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe. (B) Fractional saturation curve plotted using the EMSA data of (A). (C) Hill plot using the EMSA data from (A). (D) Semi-log saturation curve plotted using the EMSA data of (A), showing distinct breaks in the binding curve, in accordance with the multiple binding sites contained in the probe. (E) Sequence level cartoon of the fluorescently labelled probe containing r*v2827c-rv2826c* -1 to -131. (F-M) Saturation curves (F, H, J, L) and Hill plots (G, I, K, M) for each IR calculated using individual IR data gathered using mutant probes (**Fig. 4C and D, Fig. 5C and D**). For (B-D) and (F-M), points are plotted from triplicate data and display mean values with standard error of the mean.

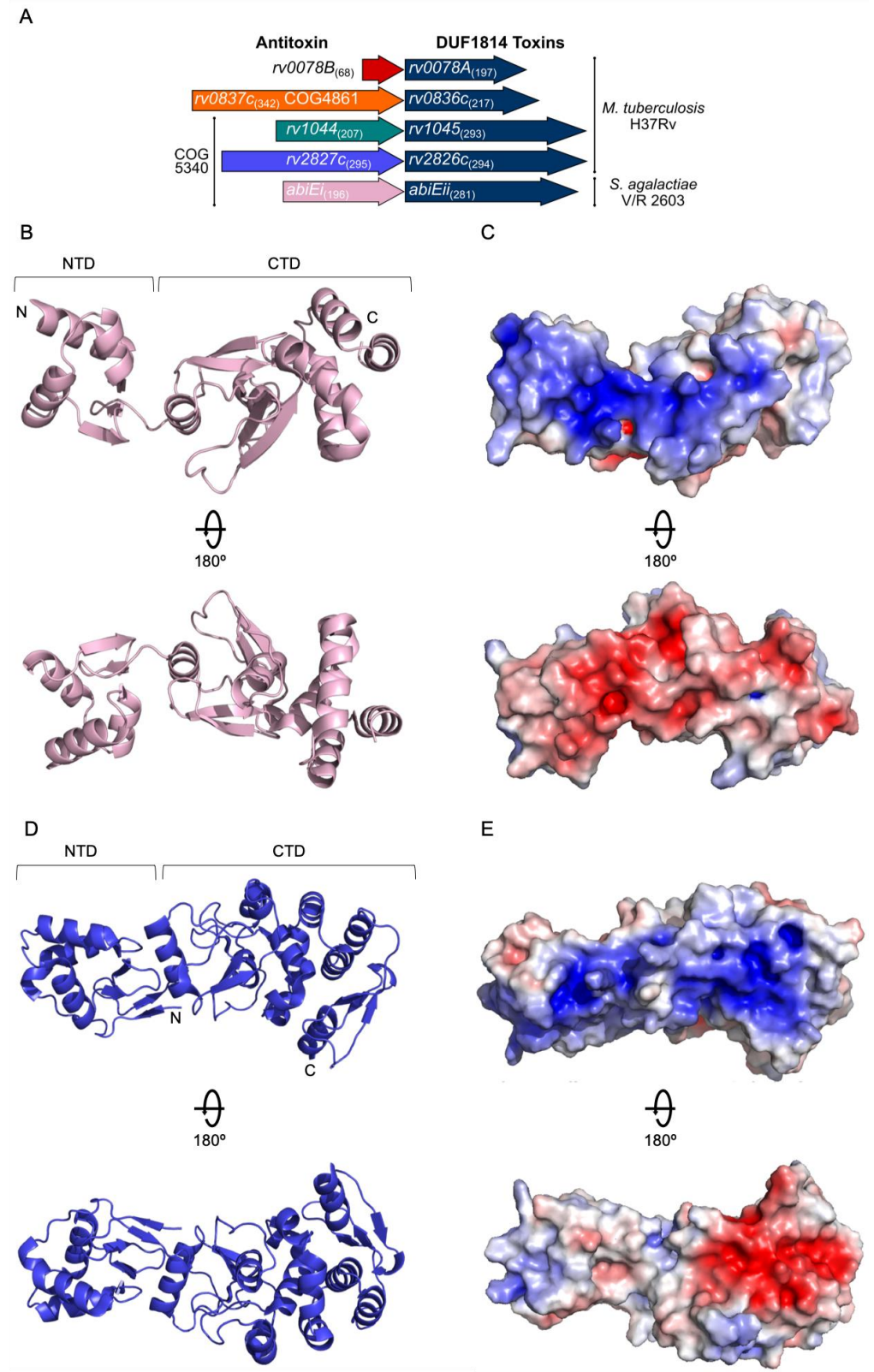
Fig. 7. Rv2827c-Rv2826c is a negatively autoregulating system in *E. coli*. (A) Promoter activity from upstream promoter regions of *abiE* (99 bp), and *rv2827c-rv2826c* and *rv1044-rv1045* (500 bp for both) detected using *lacZ* transcriptional fusions. Both the *abiE* and *rv2827c-rv2826c* constructs are active, but the *rv1044-rv1045* construct is not. Plotted data are normalized to the vector-only control. (B) Autoregulation of promoter activity by antitoxins. LacZ activity was measured from the *abiE* and *rv2827c-rv2826c* constructs with or without induction of the cognate antitoxin (AT, ±IPTG). Both AbiE and Rv2827c negatively autoregulate expression. Plotted data are normalized to the uninduced vector-only control. All data (A-C) are plotted as the means of triplicate data, and error bars show standard deviations from the mean.

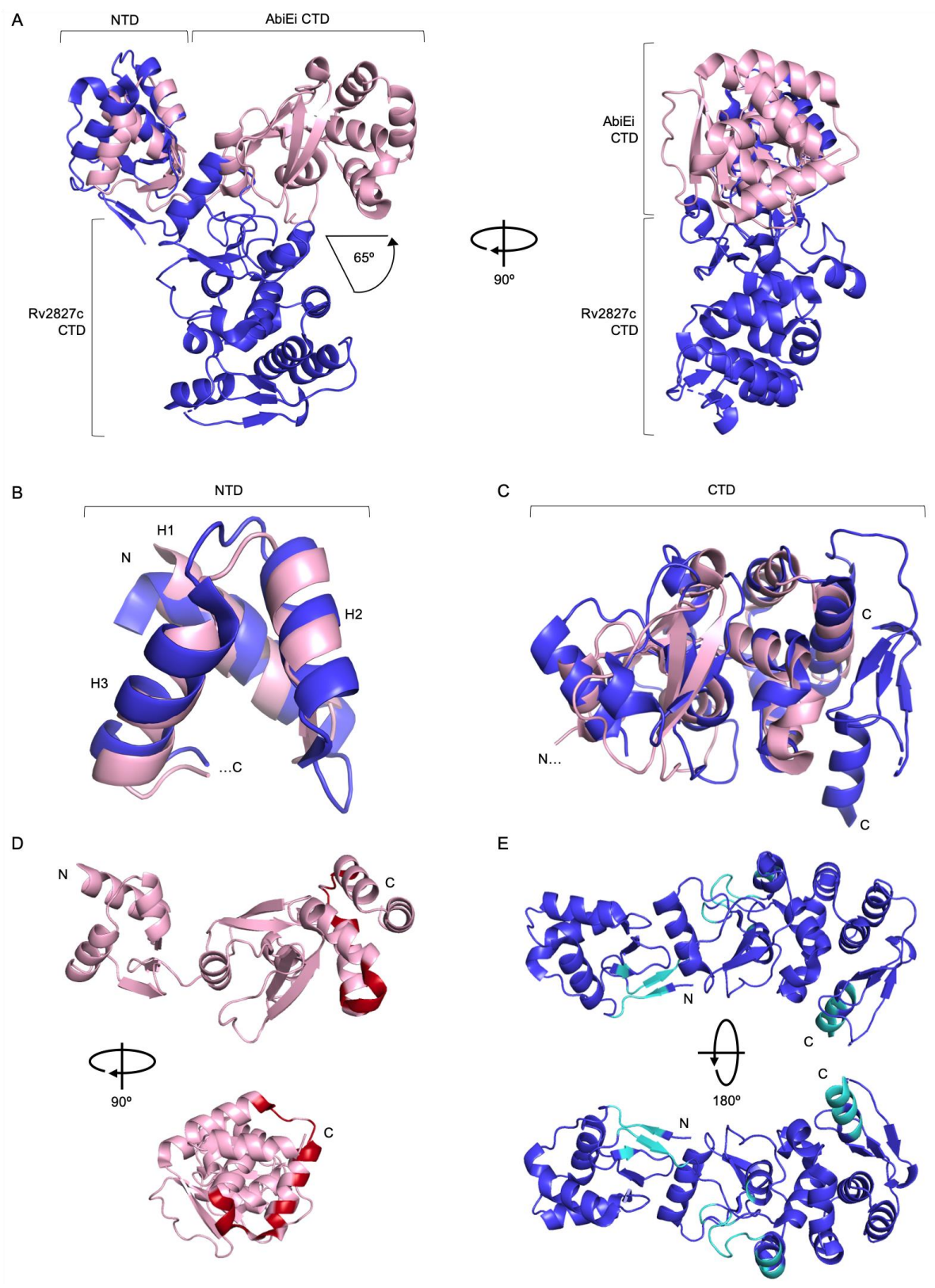
Fig. 8. Proposed model for negative autoregulation caused by Rv2827c binding to the four *rv2827c-rv2826c* promoter inverted repeats. (A) Schematic representation of the putative rv2827c-rv2826c type IV toxin-antitoxin system. Model shows both *rv2827c* and *rv2826c* being translated into the antagonistic antitoxin and toxin protein pair respectively. The antitoxin, Rv2827c has a second function and binds to the *rv2827c-rv2826c* promoter, negatively autoregulating the operon. (B) An order of binding is created by the distinct affinity values for the inverted repeats represented in the sequence level cartoon, calculated from individual IR data gathered using mutant probes (**Fig. 4C** and **D**, **Fig. 5C** and **D**). Rv2827c binds negatively co-operatively, initially to IR1 (0.0205 μM) followed by IR4 (0.121 μM), IR2 (0.862 μM) and finally IR3 (11.0 μM).

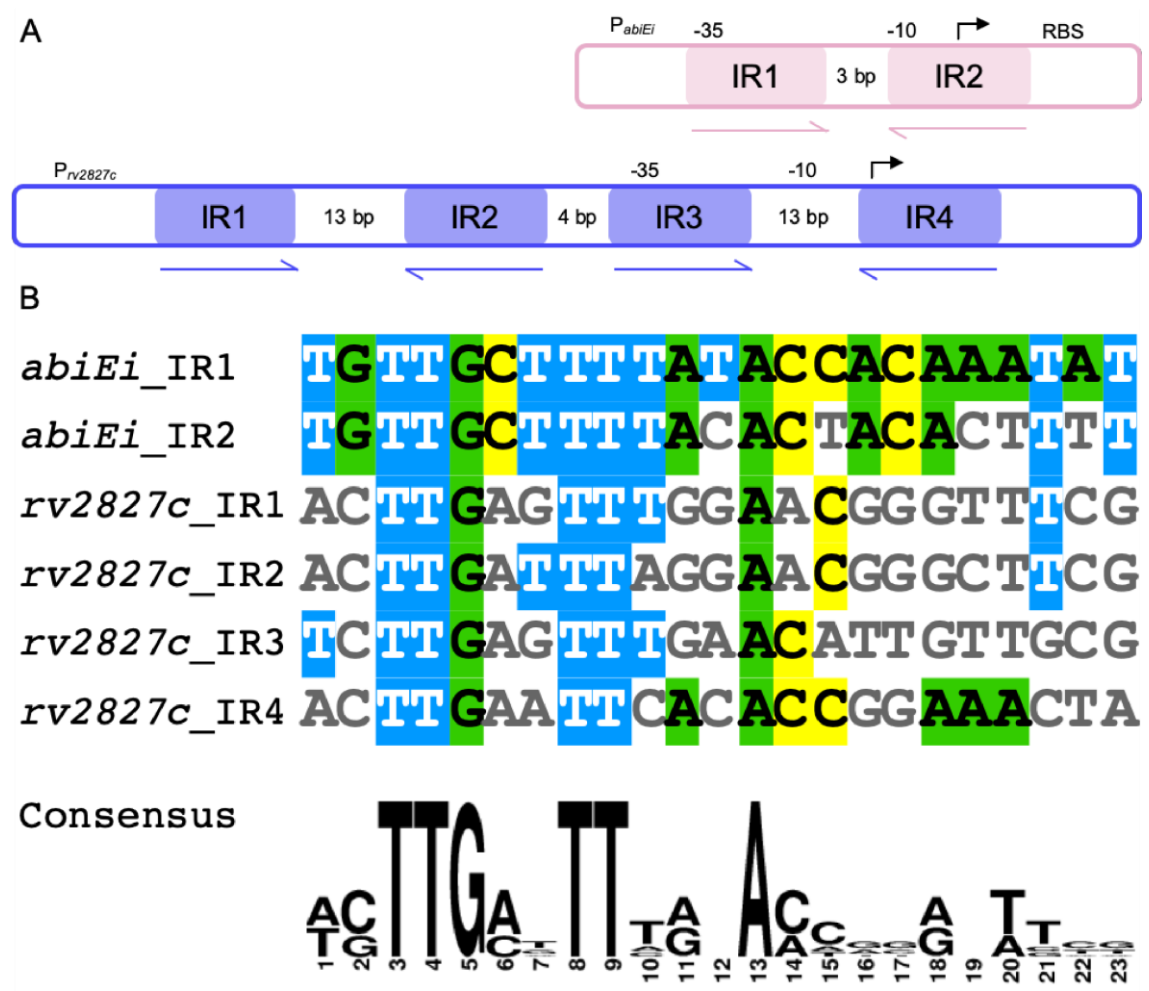
TABLES

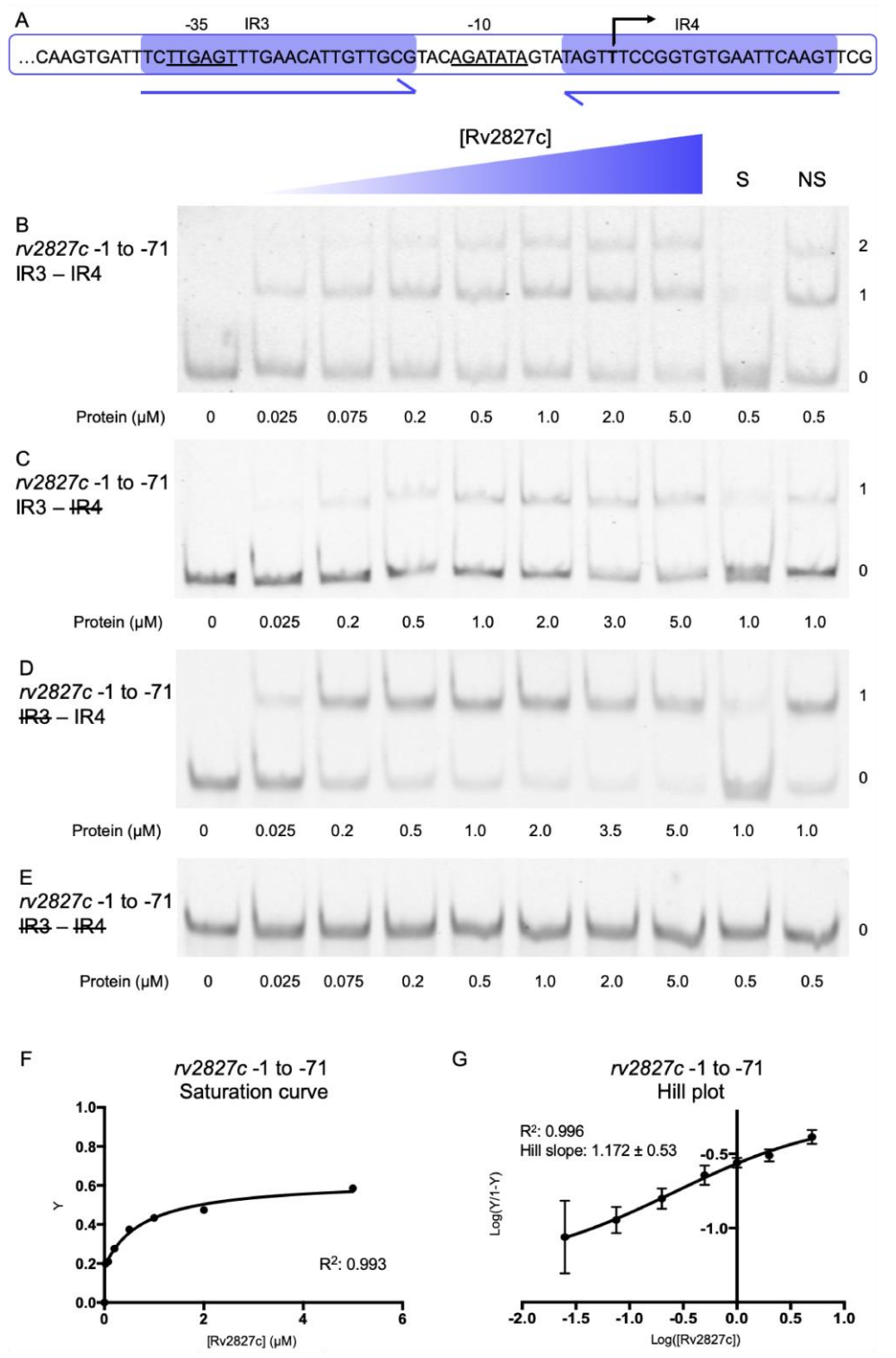
Table 1. Crystallographic Data Collection and Refinement Statistics

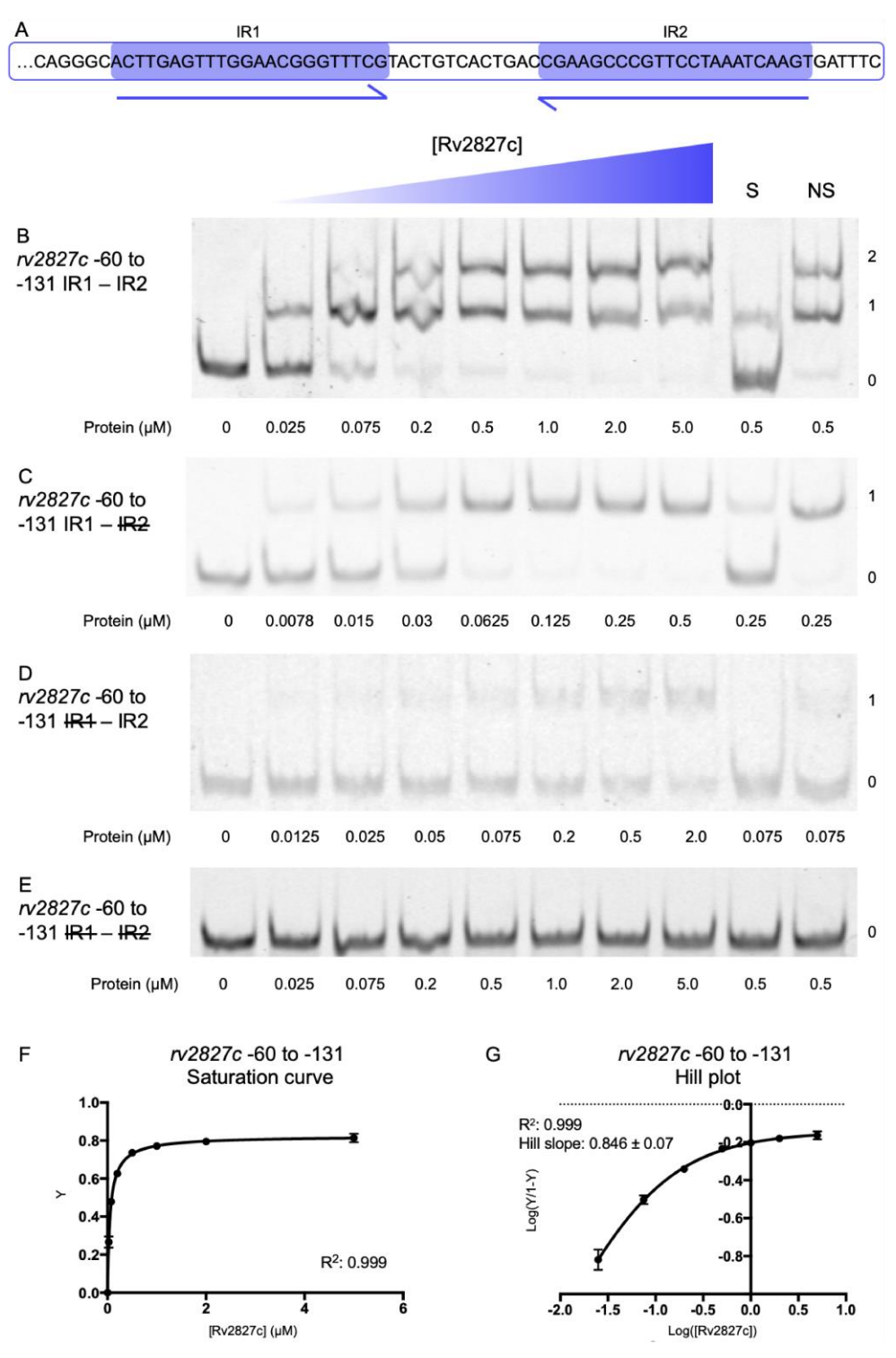
	AbiEi Native	AbiEi Se-Peak	AbiEi Se-High Remote
PDB ID Code	6Y8Q	-	-
Number of crystals	3	2	1
Beamline	Diamond I03	Diamond I03	Diamond I03
Wavelength, Å	0.9763	0.9793	0.9641
Resolution range, Å	42.11 – 1.83 (1.86 – 1.83)[a]	42.58 – 2.14 (2.19 – 2.14)	53.57 – 2.17 (2.23 – 2.17)
Space group	P1	P1	P1
Unit cell, *a b c* (Å), α β γ (°)	34.24 80.85 122.17, 102.48 96.74 100.47	34.78 81.37 122.99, 101.72 97.18 101.16	34.85 81.38 123.00, 101.74 97.31 101.19
Total reflections	207238 (10275)	443813 (13873)	129874 (8557)
Unique reflections	106620 (5213)	69714 (4469)	65917 (4312)
Multiplicity	1.9	6.4	2.0
Completeness (%)	97.4 (96.1)	99.0 (97.1)	97.9 (91.9)
Mean I/sigma(I)	7.6	6.9	6.1
R_{merge}	0.038 (0.691)	0.169 (1.036)	0.080 (0.593)
R_{meas}	0.053 (0.977)	0.184 (1.260)	0.113 (0.839)
$CC_{1/2}$	0.999 (0.471)	0.991 (0.463)	0.992 (0.599)
R_{work}	0.1812 (0.2812)	-	-
R_{free}	0.2092 (0.3100)	-	-
No. of non-hydrogen atoms	7116	-	-
Macromolecules	6397	-	-
Ligands	62	-	-
Solvent	657	-	-
Protein Residues	769	-	-
RMSD (bonds, Å)	0.012	-	-
RMSD (angles, °)	1.32	-	-
Ramachandran favored (%)	98.68	-	-
Ramachandran allowed (%)	1.32	-	-
Ramachandran outliers (%)	0.00	-	-
Average B-factor	39.61	-	-
Macromolecules	39.04	-	-
Ligands	46.01	-	-
Solvent	44.60	-	-

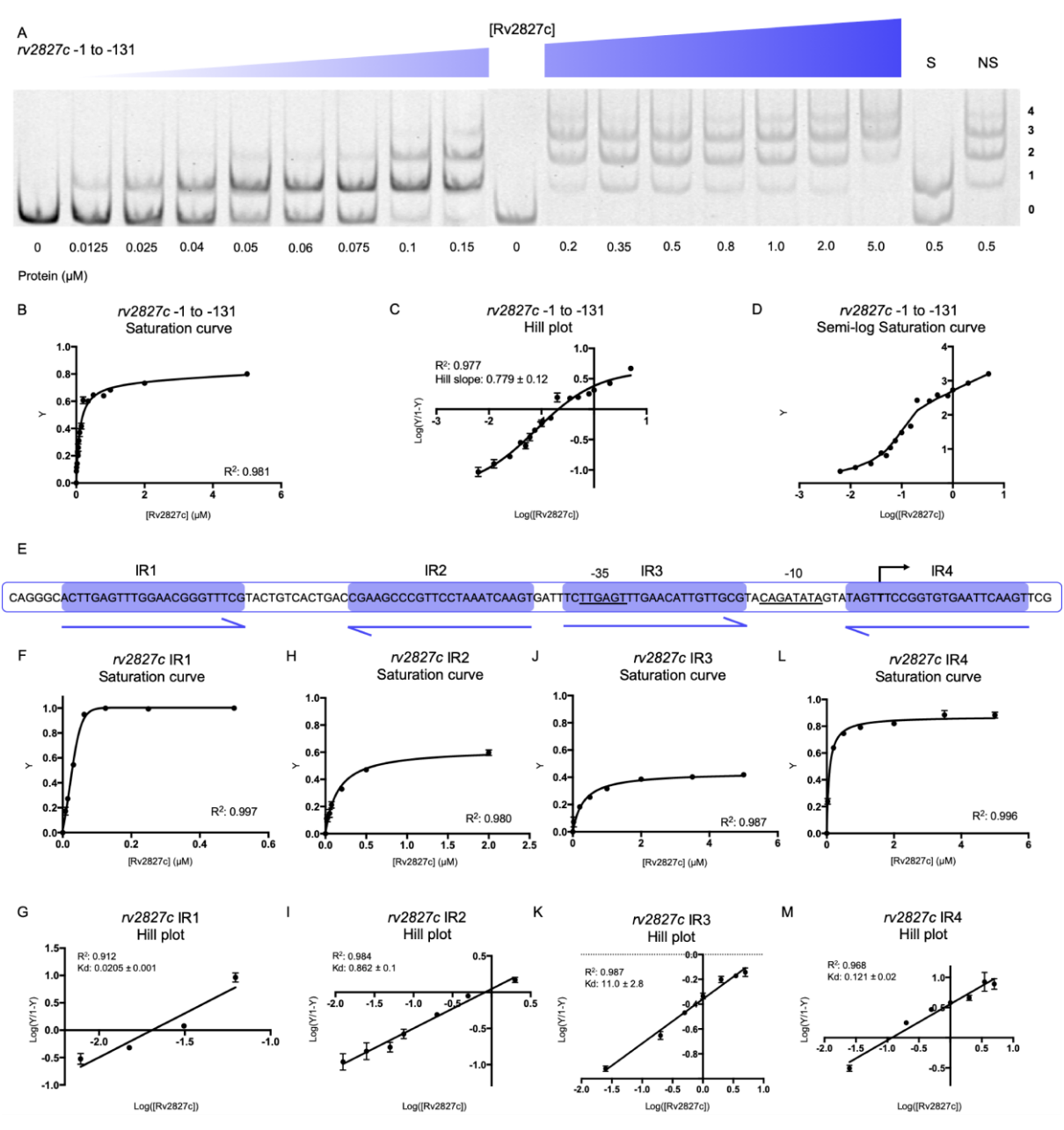

A
Antitoxin
DUF1814 Toxins
rv0078B(68)
rv0078A(197)
rv0837c(342) COG4861
rv0836c(217)
rv1044(207)
rv1045(293)
rv2827c(295)
rv2826c(294)
abiEi(196)
abiEii(281)
COG 5340
M. tuberculosis H37Rv
S. agalactiae V/R 2603
B
NTD
CTD
N
C
180º
C
180º
D
NTD
CTD
N
C
180º
E
180º

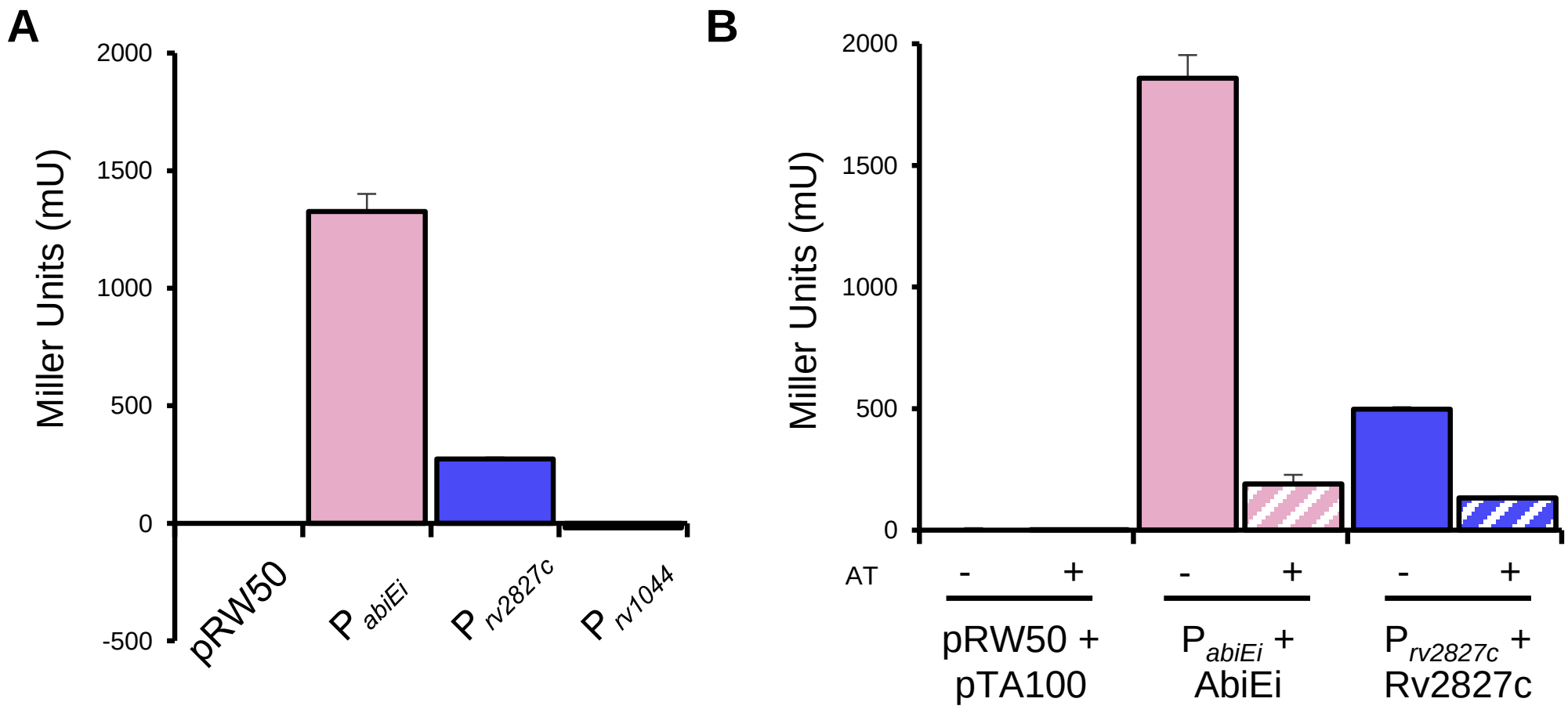

A
NTD
AbiEi CTD
65º
Rv2827c
CTD
90º
AbiEi
CTD
Rv2827c
CTD
B
NTD
H1
N
H2
H3
...C
C
CTD
C
N...
C
D
N
C
90º
C
E
N
C
180º
C
N

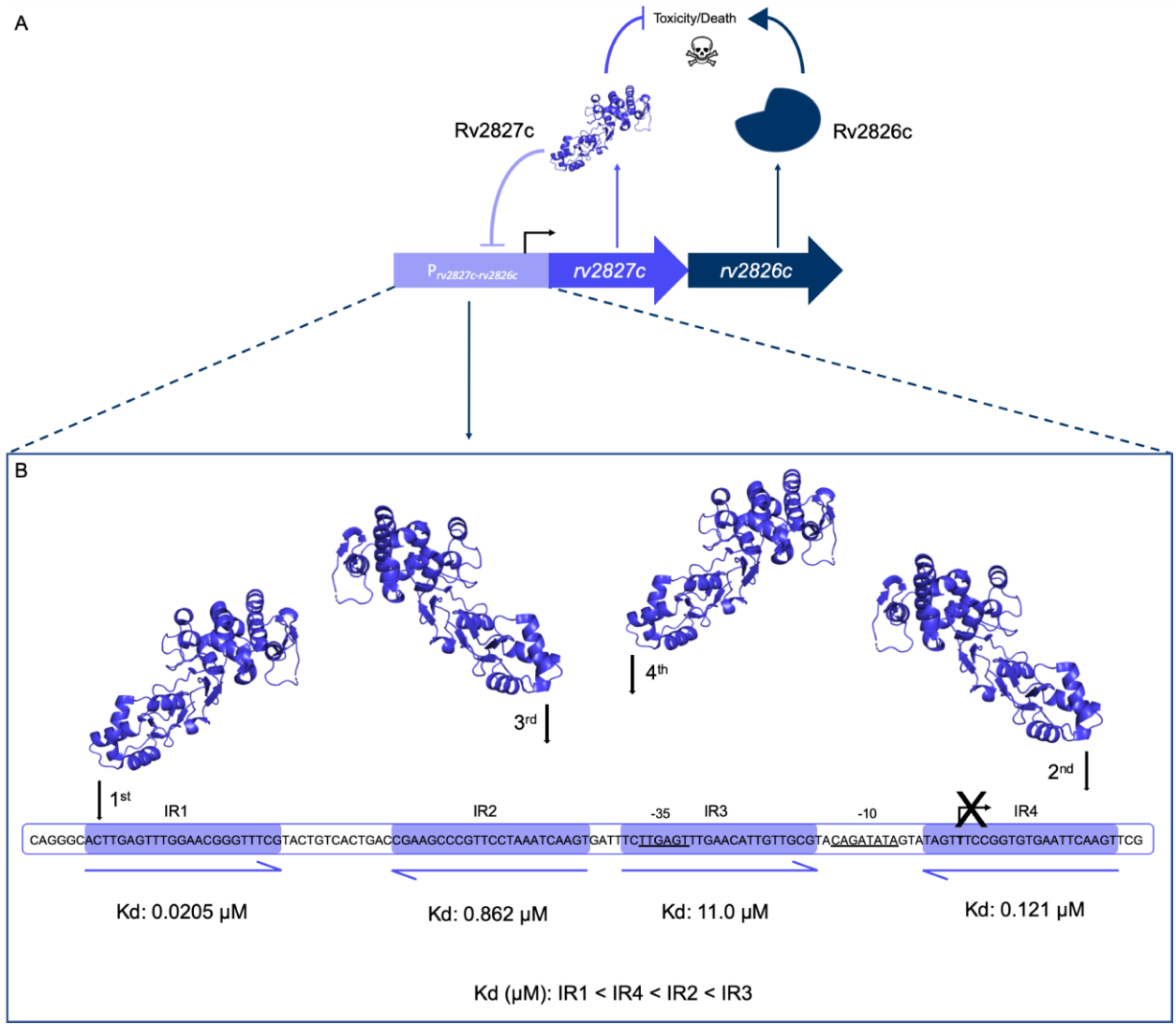

A
P abiEi
-35
-10
RBS
IR1
3 bp
IR2
P rv2827c
-35
-10
IR1
13 bp
IR2
4 bp
IR3
13 bp
IR4
B
abiEi_IR1
abiEi_IR2
rv2827c_IR1
rv2827c_IR2
rv2827c_IR3
rv2827c_IR4
Consensus


A
-35
IR3
-10
IR4
...CAAGTGATTTCTTGAGTTTGAACATTGTTGCGTACAGATATAGTATAGTTTCCGGTGTGAATTCAAGTTCG
[Rv2827c]
S
NS
B
rv2827c -1 to -71
IR3 – IR4
Protein (μM) 0 0.025 0.075 0.2 0.5 1.0 2.0 5.0 0.5 0.5
C
rv2827c -1 to -71
IR3 – ~~IR4~~
Protein (μM) 0 0.025 0.2 0.5 1.0 2.0 3.0 5.0 1.0 1.0
D
rv2827c -1 to -71
~~IR3~~ – IR4
Protein (μM) 0 0.025 0.2 0.5 1.0 2.0 3.5 5.0 1.0 1.0
E
rv2827c -1 to -71
~~IR3~~ – ~~IR4~~
Protein (μM) 0 0.025 0.075 0.2 0.5 1.0 2.0 5.0 0.5 0.5
F
rv2827c -1 to -71
Saturation curve
Y
R2: 0.993
[Rv2827c] (μM)
G
rv2827c -1 to -71
Hill plot
R2: 0.996
Hill slope: 1.172 ± 0.53
Log(Y/1-Y)
Log([Rv2827c])


A
IR1
IR2
…CAGGGCACTTGAGTTTGGAACGGGTTTCGTACTGTCACTGACCGAAGCCCGTTCCTAAATCAAGTGATTTC
[Rv2827c]
S
NS
B
rv2827c -60 to -131 IR1 – IR2
2
1
0
Protein (μM) 0 0.025 0.075 0.2 0.5 1.0 2.0 5.0 0.5 0.5
C
rv2827c -60 to -131 IR1 – IR2
1
0
Protein (μM) 0 0.0078 0.015 0.03 0.0625 0.125 0.25 0.5 0.25 0.25
D
rv2827c -60 to -131 IR1 – IR2
1
0
Protein (μM) 0 0.0125 0.025 0.05 0.075 0.2 0.5 2.0 0.075 0.075
E
rv2827c -60 to -131 IR1 – IR2
0
Protein (μM) 0 0.025 0.075 0.2 0.5 1.0 2.0 5.0 0.5 0.5
F
rv2827c -60 to -131
Saturation curve
Y
1.0
0.8
0.6
0.4
0.2
0.0
R2: 0.999
0
2
4
6
[Rv2827c] (μM)
G
rv2827c -60 to -131
Hill plot
R2: 0.999
Hill slope: 0.846 ± 0.07
Log(Y/1-Y)
0.0
-0.2
-0.4
-0.6
-0.8
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
Log([Rv2827c])


A
rv2827c -1 to -131
[Rv2827c]
S
NS
4
3
2
1
0
0 0.0125 0.025 0.04 0.05 0.06 0.075 0.1 0.15 0 0.2 0.35 0.5 0.8 1.0 2.0 5.0 0.5 0.5
Protein (μM)
B
rv2827c -1 to -131
Saturation curve
R2: 0.981
[Rv2827c] (μM)
C
rv2827c -1 to -131
Hill plot
R2: 0.977
Hill slope: 0.779 ± 0.12
Log(Y/1-Y)
Log([Rv2827c])
D
rv2827c -1 to -131
Semi-log Saturation curve
Log([Rv2827c])
E
IR1
IR2
-35
IR3
-10
IR4
CAGGGCACTTGAGTTTGGAACGGGTTTCGTACTGTCACTGACCGAAGCCCGTTCCTAAATCAAGTGATTTCTTGAGTTTGAACATTGTTGCGTACAGATATAGTATAGTTTCCGGTGTGAATTCAAGTTCG
F
rv2827c IR1
Saturation curve
R2: 0.997
[Rv2827c] (μM)
H
rv2827c IR2
Saturation curve
R2: 0.980
[Rv2827c] (μM)
J
rv2827c IR3
Saturation curve
R2: 0.987
[Rv2827c] (μM)
L
rv2827c IR4
Saturation curve
R2: 0.996
[Rv2827c] (μM)
G
rv2827c IR1
Hill plot
R2: 0.912
Kd: 0.0205 ± 0.001
Log(Y/1-Y)
Log([Rv2827c])
I
rv2827c IR2
Hill plot
R2: 0.984
Kd: 0.862 ± 0.1
Log(Y/1-Y)
Log([Rv2827c])
K
rv2827c IR3
Hill plot
R2: 0.987
Kd: 11.0 ± 2.8
Log(Y/1-Y)
Log([Rv2827c])
M
rv2827c IR4
Hill plot
R2: 0.968
Kd: 0.121 ± 0.02
Log(Y/1-Y)
Log([Rv2827c])


A
Miller Units (mU)
2000
1500
1000
500
0
-500
pRW50
P abiEi
P rv2827c
P rv1044
B
Miller Units (mU)
2000
1500
1000
500
0
AT
-
+
-
+
-
+
pRW50 + pTA100
P abiEi + AbiEi
P rv2827c + Rv2827c


A
Toxicity/Death
Rv2827c
Rv2826c
P rv2827c-rv2826c
rv2827c
rv2826c
B
1st
2nd
3rd
4th
IR1
IR2
IR3
IR4
-35
-10
CAGGGCACTTGAGTTTGGAACGGGTTTCGTACTGTCACTGACCGAAGCCCGTTCCTAAATCAAGTGATTTCTTGAGTTTGAACATTGTTGCGTACAGATATAGTATAGTTTCCGGTGTGAATTCAAGTTCG
Kd: 0.0205 μM
Kd: 0.862 μM
Kd: 11.0 μM
Kd: 0.121 μM
Kd (μM): IR1 < IR4 < IR2 < IR3

SUPPLEMENTARY DATA

Antitoxin autoregulation of *M. tuberculosis* toxin-antitoxin expression through negative cooperativity arising from multiple inverted repeat sequences

Izaak N. Beck, Ben Usher, Hannah G. Hampton, Peter C. Fineran, Tim R. Blower*

SUPPLEMENTARY FIGURES

Fig. S1. AbiEi and Rv2827c have similar folds. Structure-based sequence alignment of AbiEi and Rv2827c, drawn by hand using output from PROMALS3D. Blue arrows represent β-sheets and pink ovals represent α-helices; numbers indicate amino acid positions.

Fig. S2. AbiEi binds with positive co-operativity to the IR1-IR2 region of the *abiE* promoter. (A) Sequence level cartoon of the fluorescently labelled probe containing IR1-IR2, with -35, -10, transcriptional start and ribosome binding site (RBS) indicated. (B) Electrophoretic mobility shift assay (EMSA) of titrated AbiEi with the probe in (A). (C) EMSA of titrated AbiEi with the probe in (A) altered by replacing IR2 with polyC. (D) EMSA of titrated AbiEi with the probe in (A) altered by replacing IR1 with polyC. (E) EMSA of titrated AbiEi with the probe in (A) altered by replacing both IR1 and IR2 with polyC. For (B-E); protein concentrations are shown on each panel together with the binding events (0, 1 or 2); S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe; numbering -1 to -71 denotes the promoter region included in the probe, upstream of the translational start site in order to include all of IR2. (F) Fractional saturation curve plotted using the EMSA data of (B). (G) Hill plot using the EMSA data from (B). For (F) and (G), points are plotted from triplicate data and display mean values with standard error of the mean.

Fig. S3. AbiEi and Rv2827c do not bind non-cognate promoters. (A) Electrophoretic mobility shift assay (EMSA) of titrated AbiEi with *rv2827c-rv2826c* promoter -1 to -71. (B) EMSA of titrated AbiEi with *rv2827c-rv2826c* promoter -61 to -131. (C) EMSA of titrated Rv2827c with *abiE* promoter -1 to -71. For (A-C); protein concentrations are shown below (C) together with the binding events (0, 1 or 2); S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe; numbering denotes the promoter region included in the probe, upstream of the translational start site in order to include all of the respective IR sequences.

Fig. S4. Rv1044 does not bind the cognate promoter but is capable of DNA-binding. (A) EMSA of titrated Rv1044 with *rv1044-rv1045* promoter -1 to -71. (B) EMSA of titrated Rv1044 with *rv1044-rv1045* promoter -61 to -131. (C) EMSA of titrated Rv1044 with *rv2827c-rv2826c* promoter -1 to -71. (D) EMSA of titrated Rv1044 with *rv2827c-rv2826c* promoter -61 to -131. (E) EMSA of titrated Rv1044 with *abiE* promoter -1 to -71. For (A-E); protein concentrations are shown on each panel together with the binding events (0, 1 or 2); S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe; numbering denotes the promoter region included in the probe, upstream of the translational start site in order to include all of the respective IR sequences. (F) Fractional saturation curve plotted using the EMSA data of (E). (G) Hill plot using the EMSA data from (E). For (F) and (G), points are plotted from triplicate data and display mean values with standard error of the mean.

SUPPLEMENTARY TABLES

Table S1. Oligonucleotides used in this study

Primer	Sequence[a]	Notes (Organism/Gene)
pRW50 cloning		
TRB1072	TTTGAATTCGATTTTGTTATCACAATAAATTGAGG	FWD EcoRI, 99 bp upstream of *abiEi*, *S. agalactiae*
TRB1047	TTTAAGCTTTACGGCCCCCCACTTGTTGC	REV HindIII, 99 bp upstream of *abiEi*, *S. agalactiae*
TRB1042	TTTGAATTCGCCCAAGCATCGGCTGGC	FWD EcoRI, 500 bp upstream of *rv2827c*, *M. tuberculosis*
TRB1043	TTTAAGCTTCCGAACTTGAATTCACACCGG	REV HindIII, 500 bp upstream of *rv2827c*, *M. tuberculosis*
TRB1040	TTTGAATTCGGGTCCCAACCGAGCGGC	FWD EcoRI, 500 bp upstream of *rv1044*, *M. tuberculosis*
TRB1041	TTTAAGCTTATTAGGTGATGGAGGCCAAGGCC	REV HindIII, 500 bp upstream of *rv1044*, *M. tuberculosis*
pSAT1-LIC cloning		
TRB873	TTAATGCAGCTGATTAATACG	FWD pSAT LIC sequencing
TRB875	TACTCAAGCTTATGCATGC	REV pSAT LIC sequencing
TRB1048	CAACAGCAGACGGGAGGTTCAAAAAAAGAGATTCTACTCGATTTTATAG	FWD *abiEi* LIC, *S. agalactiae*
TRB1049	GCGAGAACCAAGGAAAGGTTATTATATTAGAACCTCCAGAGTTTGTTTAAC	REV *abiEi* LIC, *S. agalactiae*
TRB1022	CAACAGCAGACGGGAGGTGTGAGCCCAGCCGGCGCC	FWD *rv2827c* LIC, *M. tuberculosis*
TRB1023	GCGAGAACCAAGGAAAGGTTATTACGCCTTGCCGATCACGCGCAGC	REV *rv2827c* LIC, *M. tuberculosis*
TRB1018	CAACAGCAGACGGGAGGTTGTGCAAAACCGTATCTAATTGATACGATTGCGC	FWD *rv1044* LIC, *M. tuberculosis*
TRB1019	GCGAGAACCAAGGAAAGGTTATTACGCCGATGCTCGCTTCGG	REV *rv1044* LIC, *M. tuberculosis*

pTA100 cloning

TRB1052	TTTGAATTCAGGAGGACAGGGATGTCAAAAAA AGAGATTCTACTC	FWD EcoRI, *abiEi*, *S. agalactiae*
TRB1053	TTTAAGCTTGGTTATTATATTAGAACCTCCAGA GTTTG	REV HindIII, *abiEi*, *S. agalactiae*
PF1334	TTTCATATGCAATTGAGGAGGACAGGGATGGT GAGCCCAGCCG	FWD NdeI/MfeI, *rv2827c, M. tuberculosis*
PF1335	TTTACTAGTCCCGGGGTCACGCCTTGCCGATC	REV SpeI/XmaI, *rv2827c, M. tuberculosis*
PF1330	TTTCATATGCAATTGAGGAGGACAGGGATGTG TGCAAAACCGTATCTAA	FWD NdeI/MfeI, *rv1044, M. tuberculosis*
PF1331	TTTACTAGTCCCGGGCTTGGTCACGCCGATG	REV SpeI/XmaI, *rv1044, M. tuberculosis*

EMSA probe primers and templates

TRB1067	TGCGCACTGACAAAAGCTT	REV EMSA untagged
TRB1068	/56-FAM/TGCGCACTGACAAAAGCTT	REV EMSA 56-FAM (fluorescein) tagged
TRB1061	AAAAGAAAATGTTGCTTTTATACCACAAATATT GTAAAATTGTAGTGTAAAAGCAACAAGTGGGG GGCCGTAAGCTTTTGTCAGTGCGCA	*S. agalactiae / abiEi -1 to -71* WT (**Fig. S2B**, **Fig. S3C**, **Fig. S4E**)
TRB1065	AAAAGAAAATGTTGCTTTTATACCACA	FWD for TRB1061, TRB1063, *S. agalactiae, abiEi*
TRB1062	AAAAGAAAACCCCCCCCCCCTACCACAAATATT GTAAAATTGTAGTGTAAAAGCAACAAGTGGGG GGCCGTAAGCTTTTGTCAGTGCGCA	*S. agalactiae / abiEi -1 to -71* Mutant; inverted repeat 1 poly-C track substitution (**Fig. S2D**)
TRB1066	AAAAGAAAACCCCCCCCC	FWD for TRB1062, TRB1064, *S. agalactiae, abiEi*
TRB1063	AAAAGAAAATGTTGCTTTTATACCACAAATATT GTAAAATTGTAGTGCCCCCCCCCCCAGTGGGG GGCCGTAAGCTTTTGTCAGTGCGCA	*S. agalactiae / abiEi -1 to -71* Mutant; inverted repeat 2 poly-C track substitution (**Fig. S2C**)
TRB1064	AAAAGAAAACCCCCCCCCCCTACCACAAATATT GTAAAATTGTAGTGCCCCCCCCCCCAGTGGGG GGCCGTAAGCTTTTGTCAGTGCGCA	*S. agalactiae / abiEi -1 to -71* Mutant; inverted repeat 1 & 2 poly-C track substitution (**Fig. S2E**)

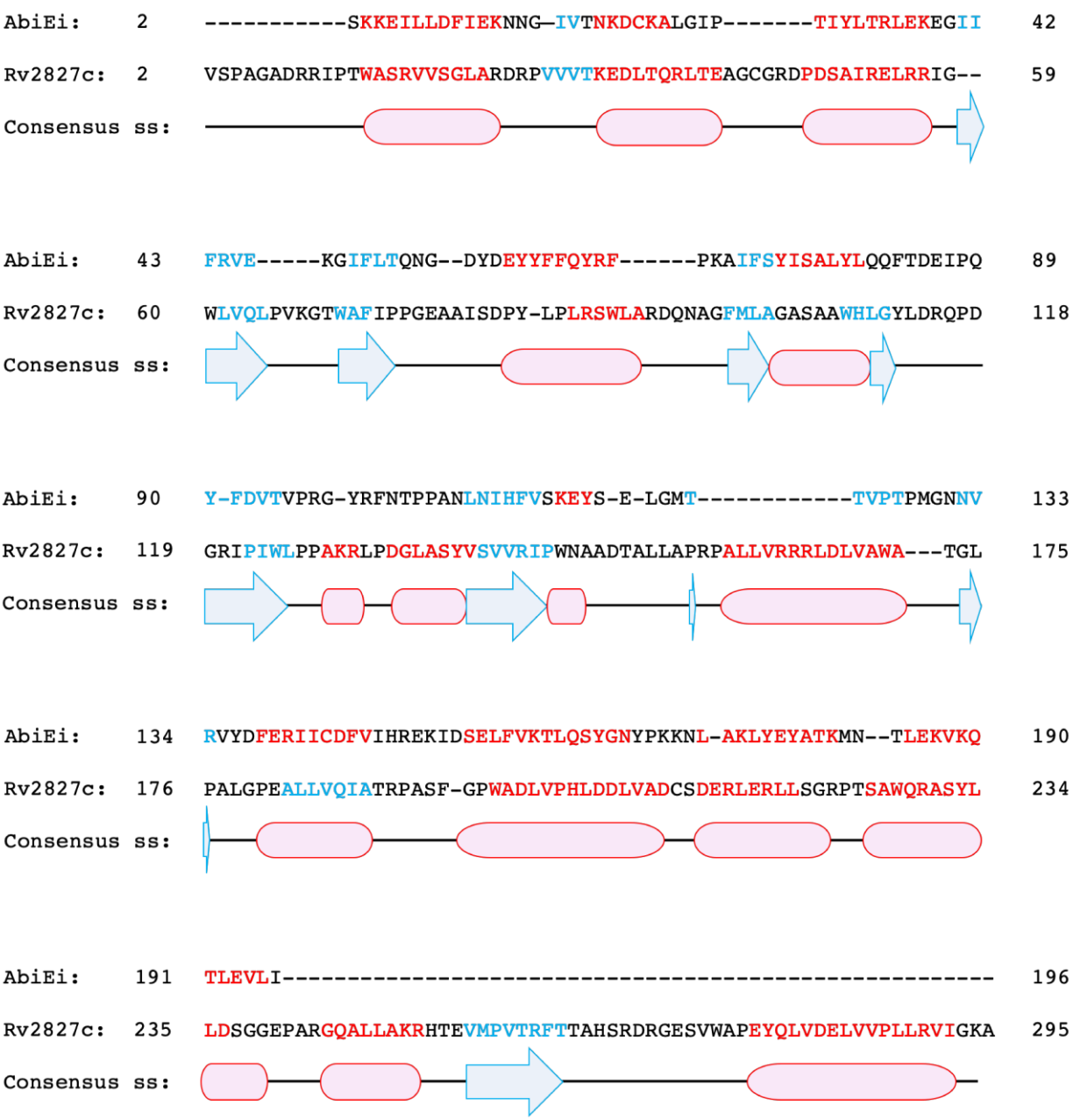
TRB1086	AACTAGGCGCGCCTAGCCTGGACGAGTCCCCG GGCCGACATTCGCCCGAGGCCTTGGCCTCCAT CACCTAAAAGCTTTTGTCAGTGCGCA	*M. tuberculosis* H37Rv */ rv1044, -1 to -71* WT (**Fig. S4A**)
TRB1087	AACTAGGCGCGCCTAG	FWD for TRB1086, *M. tuberculosis, rv1044*
TRB1102	GTATCTGCGACAAGGGCAGCGTCGATGCCTCG ACATGCAGAGTCGGTGTTCGCTTCACGCGAAC TAGGCGCAAGCTTTTGTCAGTGCGCA	*M. tuberculosis* H37Rv */ rv1044, -61 to -131* WT (**Fig. S4B**)
TRB1103	GTATCTGCGACAAGGGCAG	FWD for TRB1102, *M. tuberculosis, rv1044*
TRB1104	CAAGTGATTTCTTGAGTTTGAACATTGTTGCGT ACAGATATAGTATAGTTTCCGGTGTGAATTCAA GTTCGAAGCTTTTGTCAGTGCGCA	*M. tuberculosis* H37Rv */ rv2827c, -1 to -71* WT (**Fig. 4B**, **Fig. S3A, Fig. S4C**)
TRB1105	CAAGTGATTTCTTGAGTTTGAACATTG	FWD for TRB1104, TRB1271, *M. tuberculosis, rv2827c*
TRB1106	CAGGGCACTTGAGTTTGGAACGGGTTTCGTAC TGTCACTGACCGAAGCCCGTTCCTAAATCAAGT GATTTCAAGCTTTTGTCAGTGCGCA	*M. tuberculosis* H37Rv */ rv2827c, -61 to -131* WT (**Fig. 5B, Fig. S3B, Fig. S4D**)
TRB1107	CAGGGCACTTGAGTTTGGAAC	FWD for TRB1106, TRB1277, full-length *-1 to -131* (**Fig. 6A**), *M. tuberculosis, rv2827c*
TRB1108	TGGCATTCAATCGATGGCTTCCTAGTTTTAGAT GATTAGGGCTTGTCCCAAATGGATTGAGAGGT TGACAAAGCTTTTGTCAGTGCGCA	Plasmid pEFER, 12851-12920 bp NS probe (**Fig. 4 – 6, Fig. S2 – S4**)
TRB1109	TGGCATTCAATCGATGGCTT	FWD plasmid pEFER NS probe
TRB1271	CAAGTGATTCCCCCCCCCCCCCCCCCCCCCTA CAGATATAGTATAGTTTCCGGTGTGAATTCAAG TTCGAAGCTTTTGTCAGTGCGCA	*M. tuberculosis* H37Rv */ rv2827c, -1 to -71* Mutant; inverted repeat 3 poly-C track substitution (**Fig. 4D**)
TRB1272	CAAGTGATTCCCCCCCCC	FWD for TRB1271, TRB1274, *M. tuberculosis, rv2827c*
TRB1273	CAAGTGATTTCTTGAGTTTGAACATTGTTGCGT ACAGATATAGTACCCCCCCCCCCCCCCCCCCC CCTCGAAGCTTTTGTCAGTGCGCA	*M. tuberculosis* H37Rv */ rv2827c, -1 to -71* Mutant; inverted repeat 4 poly-C track substitution (**Fig. 4C**)
TRB1274	CAAGTGATTCCCCCCCCCCCCCCCCCCCCCCTA CAGATATAGTACCCCCCCCCCCCCCCCCCCC CTCGAAGCTTTTGTCAGTGCGCA	*M. tuberculosis* H37Rv */ rv2827c, -1 to -71* Mutant; inverted repeat 3 & 4 poly-C track substitution (**Fig. 4E**)

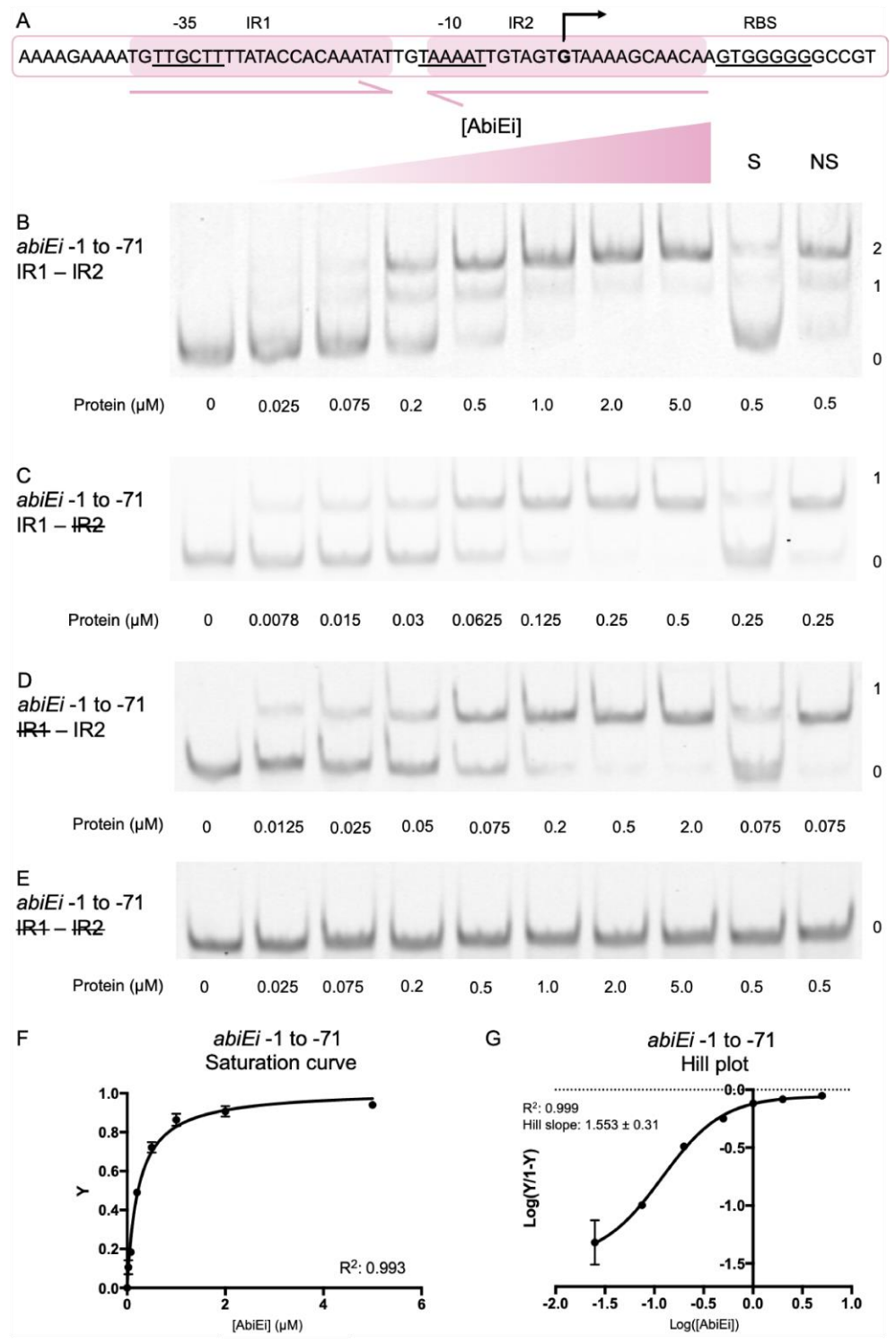
TRB1275	CAGGGCCCCCCCCCCCCCCCCCCCCCCCTACT GTCACTGACCGAAGCCCGTTCCTAAATCAAGT GATTTAAGCTTTTGTCAGTGCGCA	*M. tuberculosis* H37Rv */ rv2827c, -61 to -131* Mutant; inverted repeat 1 poly-C track substitution (**Fig. 5D**)
TRB1276	CAGGGCCCCCCCCCC	FWD for TRB1275, TRB1278, *M. tuberculosis, rv2827c*
TRB1277	CAGGGCACTTGAGTTTGGAACGGGTTTCGTAC TGTCACTGACCCCCCCCCCCCCCCCCCCCCCC GATTTAAGCTTTTGTCAGTGCGCA	*M. tuberculosis* H37Rv */ rv2827c, -61 to -131* Mutant; inverted repeat 2 poly-C track substitution (**Fig. 5C**)
TRB1278	CAGGGCCCCCCCCCCCCCCCCCCCCCCCTACT GTCACTGACCCCCCCCCCCCCCCCCCCCCCG ATTTAAGCTTTTGTCAGTGCGCA	*M. tuberculosis* H37Rv */ rv2827c, -61 to -131* Mutant; inverted repeat 1 & 2 poly-C track substitution (**Fig. 5E**)

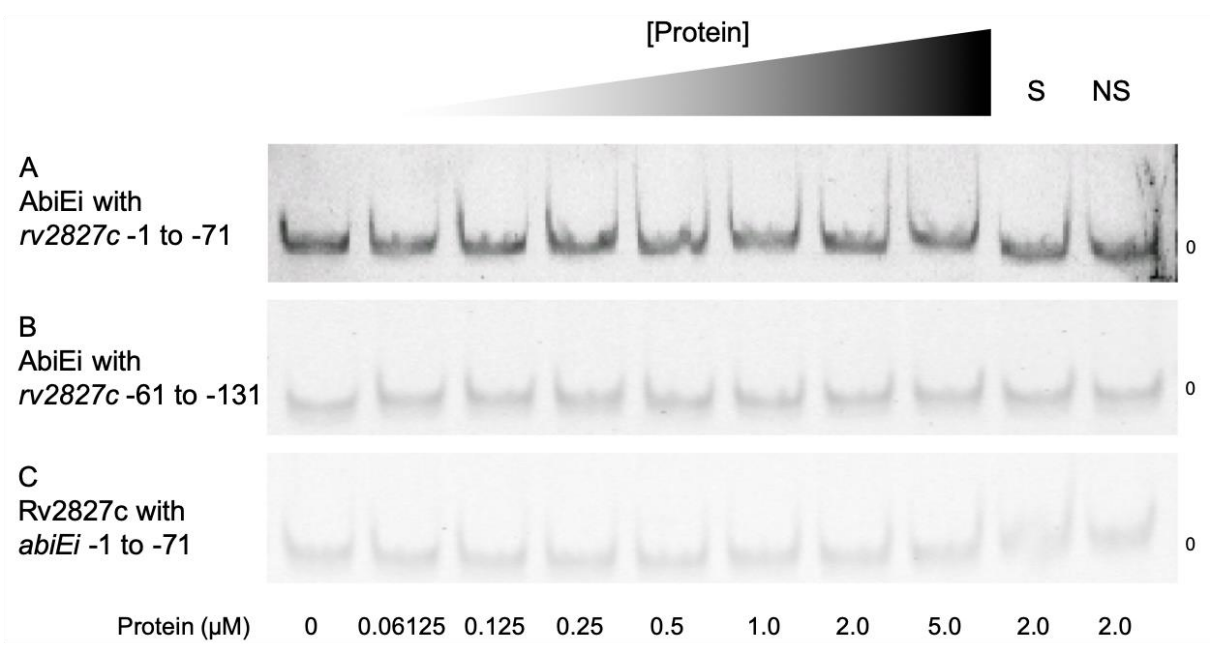
[a]EMSA probe sequences are fused with a constant region from the *lacZ* gene of pRW50, highlighted in grey. The reverse primers (TRB1067 and TRB1069) anneal to this sequence for amplification.

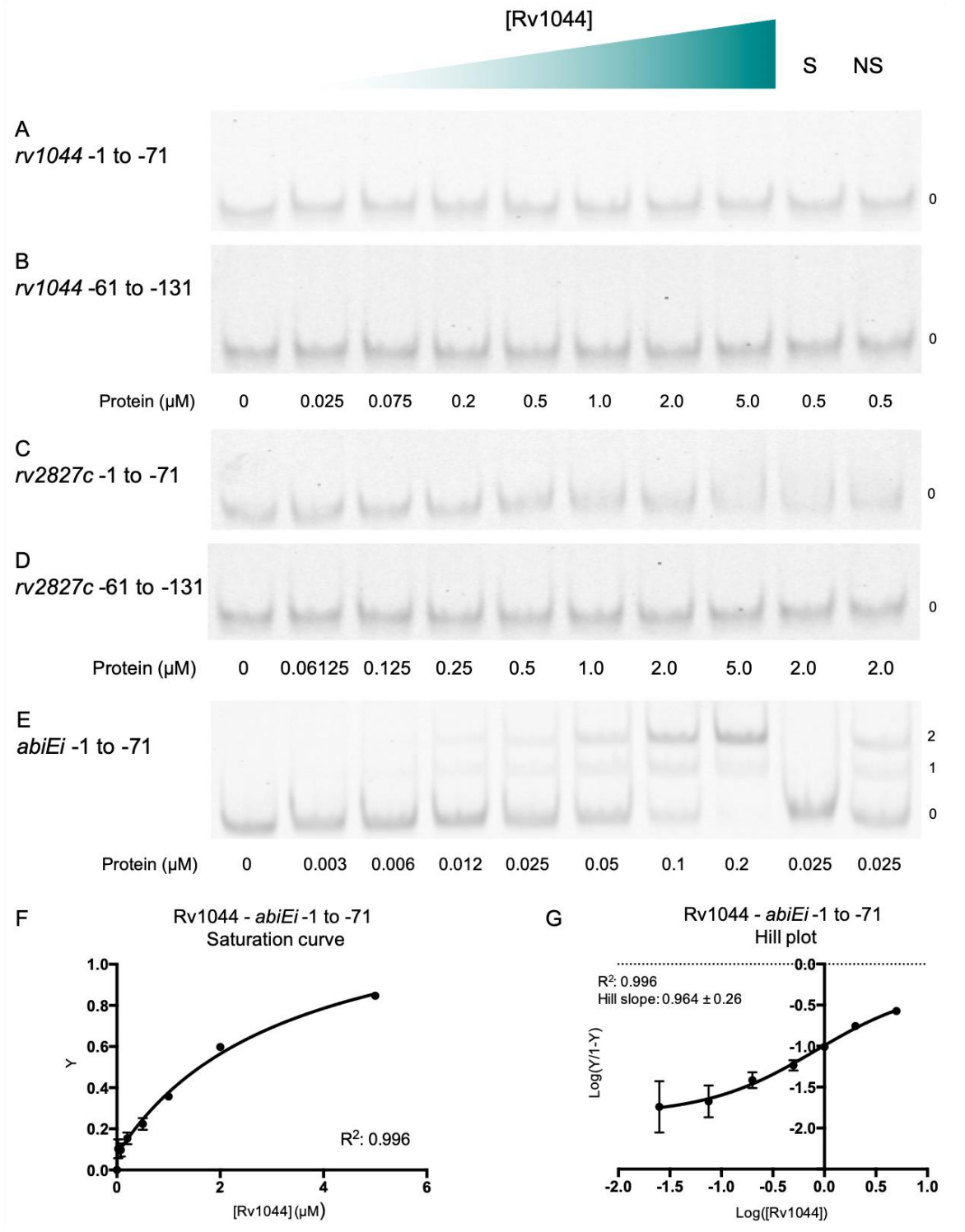
Table S2. Plasmids used in this study

Plasmid	Construct	Cloning technique	Primer set/Restriction enzymes used	Reference
pPF656	pTA100-*rv1044*	Restriction Cloning	PF1330/PF1331	This work
pPF658	pTA100-*rv2827c*	Restriction Cloning	PF1334/PF1335	This work
pRLD30	pTRB30-His_6-*abiEi*	-	-	[15]
pRW50	Tc^R	-	-	[32]
pSAT1-LIC	Ap^R	-	-	This work
pTA100	Sm^R	-	-	[5]
pTRB481	pTA100-*abiEi*	Restriction Cloning	TRB1052/TRB1053	This work
pTRB483	pRW50-500 bp upstream *rv1044*	Restriction Cloning	TRB1040/TRB1041	This work
pTRB484	pRW50-500 bp upstream *rv2827c*	Restriction Cloning	TRB1042/TRB1043	This work
pTRB486	pRW50-99 bp upstream *abiEi*	Restriction Cloning	TRB1072/TRB1047	This work
pTRB491	pSAT1-LIC-*rv1044*	Ligation Independent Cloning	TRB1018/TRB1019	This work
pTRB493	pSAT1-LIC-*rv2827c*	Ligation Independent Cloning	TRB1022/TRB1023	This work
pTRB525	pSAT1-LIC-*abiEi*	Ligation Independent Cloning	TRB1048/TRB1049	This work


AbiEi: 2 ----------SKKEILLDFIEKNNG-IVTNKDCKALGIP-------TIYLTRLEKEGII 42
Rv2827c: 2 VSPAGADRRIPTWASRVVSGLARDRPVVVTKEDLTQRLTEAGCGRDPDSAIRELRRIG-- 59
Consensus ss:
AbiEi: 43 FRVE-----KGIFLTQNG--DYDEYYFFQYRF------PKAIFSYISALYLQQFTDEIPQ 89
Rv2827c: 60 WLVQLPVKGTWAFIPPGEAAISDPY-LPLRSWLARDQNAGFMLAGASAAWHLGYLDRQPD 118
Consensus ss:
AbiEi: 90 Y-FDVTVPRG-YRFNTPPANLNIHFVSKEYS-E-LGMT----------TVPTPMGNNV 133
Rv2827c: 119 GRIPIWLPPAKRLPDGLASYVSVVRIPWNAADTALLAPRPALLVRRRLDLVAWA---TGL 175
Consensus ss:
AbiEi: 134 RVYDFERIICDFVIHREKIDSELFVKTLQSYGNYPKKNL-AKLYEYATKMN--TLEKVKQ 190
Rv2827c: 176 PALGPEALLVQIATRPASF-GPWADLVPHLDDLVADCSDERLERLLSGRPTSAWQRASYL 234
Consensus ss:
AbiEi: 191 TLEVLI---------- 196
Rv2827c: 235 LDSGGEPARGQALLAKRHTEVMPVTRFTTAHSRDRGESVWAPEYQLVDELVVPLLRVIGKA 295
Consensus ss:


A
-35
IR1
-10
IR2
RBS
AAAAGAAAATGTTGCTTTATACCACAAATATTGTAAAATTGTAGTGTAAAAGCAACAAGTGGGGGGCCGT
[AbiEi]
S
NS
B
abiEi -1 to -71
IR1 – IR2
Protein (μM) 0 0.025 0.075 0.2 0.5 1.0 2.0 5.0 0.5 0.5
C
abiEi -1 to -71
IR1 – IR2
Protein (μM) 0 0.0078 0.015 0.03 0.0625 0.125 0.25 0.5 0.25 0.25
D
abiEi -1 to -71
IR1 – IR2
Protein (μM) 0 0.0125 0.025 0.05 0.075 0.2 0.5 2.0 0.075 0.075
E
abiEi -1 to -71
IR1 – IR2
Protein (μM) 0 0.025 0.075 0.2 0.5 1.0 2.0 5.0 0.5 0.5
F
abiEi -1 to -71
Saturation curve
Y
[AbiEi] (μM)
R2: 0.993
G
abiEi -1 to -71
Hill plot
R2: 0.999
Hill slope: 1.553 ± 0.31
Log(Y/1-Y)
Log([AbiEi])


[Protein]
S
NS
A
AbiEi with
rv2827c -1 to -71
0
B
AbiEi with
rv2827c -61 to -131
0
C
Rv2827c with
abiEi -1 to -71
0
Protein (μM) 0 0.06125 0.125 0.25 0.5 1.0 2.0 5.0 2.0 2.0


[Rv1044]
S
NS
A
rv1044 -1 to -71
B
rv1044 -61 to -131
Protein (μM) 0 0.025 0.075 0.2 0.5 1.0 2.0 5.0 0.5 0.5
C
rv2827c -1 to -71
D
rv2827c -61 to -131
Protein (μM) 0 0.06125 0.125 0.25 0.5 1.0 2.0 5.0 2.0 2.0
E
abiEi -1 to -71
Protein (μM) 0 0.003 0.006 0.012 0.025 0.05 0.1 0.2 0.025 0.025
F
Rv1044 - abiEi -1 to -71
Saturation curve
Y
R2: 0.996
[Rv1044] (μM)
G
Rv1044 - abiEi -1 to -71
Hill plot
R2: 0.996
Hill slope: 0.964 ± 0.26
Log(Y/1-Y)
Log([Rv1044])